

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2001, 10:01:43 ; Search time 2123.11 Seconds

(Without alignments)  
16612.879 Million cell updates/sec

Title: US-09-215-035-1

Perfect score: 2138  
Sequence: 1 AGCAATCCGTCGCCGCC.....AAAAAAAAAAAAAAAAAAAA 2138

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2138	100.0	2138	6	ARI00763
2	2138	100.0	2138	6	ARI19934
3	2114	98.9	2114	9	HSU040434
4	1994	93.3	2129	6	E07943
5	1994	93.3	2129	6	E07943
6	1981.8	92.7	2093	9	BC003512
7	1975	92.4	2093	9	HUMMPF
8	1961	91.7	2162	9	BC009272
9	980.8	45.9	1195	9	AF180951
10	910	42.6	2135	10	D86370
11	902.8	42.2	2099	10	HS3351
12	307.6	14.4	37307	9	HS3351
13	307.6	14.4	154615	2	AC067772
14	307.6	14.4	256073	9	AE006464
15	296.6	13.9	155770	2	AC092381
16	278.2	13.0	155770	2	AC092381
17	243.2	11.4	185697	2	AC067814
18	219.8	10.3	396	6	AX093308
19	197	9.2	197	6	189984
20	117	5.3	209826	2	AC068497
21	75	3.5	75	6	189983
22	69	3.2	69	6	189982
23	54.2	2.5	2658	1	STMNRDHP
24	53.8	2.5	188123	2	AC012004
25	53.2	2.5	2490	1	AB019394
26	53.2	2.5	23990	1	SC6E10
27	52.8	2.5	89948	9	HS756523
28	52.6	2.5	10732	6	E32986
29	52.4	2.5	2884	9	HSN800512
30	52.4	2.5	3103	9	HS10N
31	52.4	2.5	3137	9	BC000235
32	52.4	2.5	3179	9	HS002389
33	52.4	2.5	3407	6	176203
34	52.4	2.5	3407	6	176203
35	52	2.4	92458	9	AL533715
36	51.6	2.4	1736	9	BC009285
37	51.4	2.4	208936	2	AC010821
38	51.2	2.4	27210	9	HS422E10
39	51.2	2.4	278911	9	AE006465
40	50.6	2.4	298166	2	AC087563
41	50.4	2.4	114793	9	AF217796
42	50.2	2.3	2000	6	108490
43	50.2	2.3	41944	3	AC005929
44	50	2.3	1263	10	MH070658
45	50	2.3	1292	10	AF009520

## ALIGNMENTS

RESULT 1  
ARI00763  
LOCUS ARI00763 2138 bp DNA  
DEFINITION Sequence 1 from patent US 6083502.  
ACCESSION ARI00763  
VERSION ARI00763.1 GI:12811561  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2138)  
AUTHORS Pastan, I. and Chang, K.  
TITLE Mesothelium antigen and methods and kits for targeting it  
JOURNAL Patent: US 6083502-A 1 04-JUL-2000;  
FEATURES  
source 1..2138  
BASE COUNT 385 a 722 c 661 g 370 t  
ORIGIN

14-FEB-2001

Query Match 100.0%; Score 2138; DB 6; Length 2138;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AGGAATTCGGTGGCGGCGACCTCCGCTCTGTGACGCGCGGACAGAGAGTACCGGT 60
DB 1 AGGAATTCGGTGGCGGCGACCTCCGCTCTGTGACGCGCGGACAGAGAGTACCGGT 60
QY 61 GGACCCAGGAGTCCCTCCCTGGGATCTACAGACCATGGCTTGCACAGCGTGCAC 120
DB 61 GGACCCAGGAGTCCCTCCCTGGGATCTACAGACCATGGCTTGCACAGCGTGCAC 120
QY 121 CCCTGTGGTGGTCTGTGGGAGCGCCCTGGACGCTCTGTCTCTGAGCTTCGGA 180
DB 121 CCCTGTGGTGGTCTGTGGGAGCGCCCTGGACGCTCTGTCTCTGAGCTTCGGA 180
QY 181 TGGGTGATCCCGGAGAGACCTGGCTGGAGAGACAGGAGAGAGTGGCCCTGGGG 240
DB 181 TGGGTGATCCCGGAGAGACCTGGCTGGAGAGACAGGAGAGAGTGGCCCTGGGG 240
QY 241 GGAGTCCGAGACCCCTCATACATTTCCAGCCCTCCCTCCGCACTCCTTGGCTTC 300
DB 241 GGAGTCCGAGACCCCTCATACATTTCCAGCCCTCCCTCCGCACTCCTTGGCTTC 300
QY 301 CCGTGTGGGAGGTGTCCGGCTGAGCAGGAGCGTGTCCGGAGAGCTGGCTGGCTTG 360
DB 301 CCGTGTGGGAGGTGTCCGGCTGAGCAGGAGCGTGTCCGGAGAGCTGGCTGGCTTG 360
QY 361 GCACAGAGAAATGTCAAGCTCTCAACAGAGAGCTGGCTGTGAGTCCGCTCTCT 420
DB 361 GCACAGAGAAATGTCAAGCTCTCAACAGAGAGCTGGCTGTGAGTCCGCTCTCT 420
QY 421 GAGCCCCCGAGAGACCTGAGGCGCTCCATTTGAGCGTGTGATTTCTCAACCAAT 480
DB 421 GAGCCCCCGAGAGACCTGAGGCGCTCCATTTGAGCGTGTGATTTCTCAACCAAT 480
QY 481 GCGTGTTCGGGGGCGCCAGGCGCTTGACCCGTTTCTTCCGCAATACAGAGCCAAATG 540
DB 481 GCGTGTTCGGGGGCGCCAGGCGCTTGACCCGTTTCTTCCGCAATACAGAGCCAAATG 540
QY 541 GACCTGCTCCCGAGGAGGCGCTCCGAGGAGACAGCGGCTGTGCTGCGGCTGTGCTG 600
DB 541 GACCTGCTCCCGAGGAGGCGCTCCGAGGAGACAGCGGCTGTGCTGCGGCTGTGCTG 600
QY 601 TGGGTGTGGGGGCTCTCTGATGAGGAGGCTGATGTGCGGCTCTGGAGGCGTGGCT 660
DB 601 TGGGTGTGGGGGCTCTCTGATGAGGAGGCTGATGTGCGGCTCTGGAGGCGTGGCT 660
QY 661 TGGGACCTGCTGGGCGCTTGTGTGGCGAGTGTGCGGCGAGTGTGCTACCCGGCTGGTG 720
DB 661 TGGGACCTGCTGGGCGCTTGTGTGGCGAGTGTGCGGCGAGTGTGCTACCCGGCTGGTG 720
QY 721 AGCTGCCCGGAGACCTTGAGACAGAGAGAGGAGCGGCGCTCTGTGAGGCG 780
DB 721 AGCTGCCCGGAGACCTTGAGACAGAGAGAGGAGCGGCGCTCTGTGAGGCG 780
QY 781 GGGGAGACCCCTTACGGGCGCGCTGAGATGTGTCTCTCAGATGAGAGCTGTGGG 840
DB 781 GGGGAGACCCCTTACGGGCGCGCTGAGATGTGTCTCTCAGATGAGAGCTGTGGG 840
QY 841 GGGCTGTGACCTGTGGGCGAGCCATCATCTGCGAGATCCCGGAGGAGTGTGGCC 900
DB 841 GGGCTGTGACCTGTGGGCGAGCCATCATCTGCGAGATCCCGGAGGAGTGTGGCC 900
QY 901 GCGTGGGCGGAGCTCTCTGCGGAGCCATCTGCGGAGCGCTGAGACGAGCATCTTC 960
DB 901 GCGTGGGCGGAGCTCTCTGCGGAGCCATCTGCGGAGCGCTGAGACGAGCATCTTC 960
QY 961 CGGCGCGGCTTCCGGGAGGAGTGGAGAGACGCTGTCTTCAAGGCAAGAGCGCGC 1020
DB 961 CGGCGCGGCTTCCGGGAGGAGTGGAGAGACGCTGTCTTCAAGGCAAGAGCGCGC 1020
```

```
QY 1021 GAGATAGACAGAGACCTCATCTTCTACAGAGTGGAGCTGGAAGCTCGGTGATCG 1080
DB 1021 GAGATAGACAGAGACCTCATCTTCTACAGAGTGGAGCTGGAAGCTCGGTGATCG 1080
QY 1081 GCGCTGTGGCCACCCAGATGAGCGGCTGAAGCGCATCCCTTACCTACGACAGCTG 1140
DB 1081 GCGCTGTGGCCACCCAGATGAGCGGCTGAAGCGCATCCCTTACCTACGACAGCTG 1140
QY 1141 GAGGCTCAAGACATTAAGTGGATGAGTCAACGACAGAGGTTACCCGAGTGTGATC 1200
DB 1141 GAGGCTCAAGACATTAAGTGGATGAGTCAACGACAGAGGTTACCCGAGTGTGATC 1200
QY 1201 CAGACCTGTGGGCTACCTTCTCTCAGATGAGCCCTGAGCATTTCCGAAATGATG 1260
DB 1201 CAGACCTGTGGGCTACCTTCTCTCAGATGAGCCCTGAGCATTTCCGAAATGATG 1260
QY 1261 AGCTCCCTGAGAGACCTGAGAGCTTGTGTAAGTGCAGAAAGGAGCAGAAATGATCT 1320
DB 1261 AGCTCCCTGAGAGACCTGAGAGCTTGTGTAAGTGCAGAAAGGAGCAGAAATGATCT 1320
QY 1321 CAGGCTCCTGGGGCGCCCTCCACAGGAGTGGGCAACCCGTGATGAGACCGTTTGAAGGGA 1380
DB 1321 CAGGCTCCTGGGGCGCCCTCCACAGGAGTGGGCAACCCGTGATGAGACCGTTTGAAGGGA 1380
QY 1381 AGGGGCGAGCTAGACAAAGAGACACCTAGACACCTGACCGCTTCTACCTTGGGTACTG 1440
DB 1381 AGGGGCGAGCTAGACAAAGAGACACCTAGACACCTGACCGCTTCTACCTTGGGTACTG 1440
QY 1441 TGTCTCCTCAGACCCCGAGAGCTGAGCTCGCCGCCACAGCATGTGGGCGTCAAG 1500
DB 1441 TGTCTCCTCAGACCCCGAGAGCTGAGCTCGCCGCCACAGCATGTGGGCGTCAAG 1500
QY 1501 CCGGAGAGCTGAGACAGTGTGACCCAGAGGAGCGAGTGGAGTCTCTATCCCAAGGCGCG 1560
DB 1501 CCGGAGAGCTGAGACAGTGTGACCCAGAGGAGCGAGTGGAGTCTCTATCCCAAGGCGCG 1560
QY 1561 CTTCCTTCCAGAACATGAGAGGAGTCCGATCTTCTGAGATCCAGTCTCTCTGGGT 1620
DB 1561 CTTCCTTCCAGAACATGAGAGGAGTCCGATCTTCTGAGATCCAGTCTCTCTGGGT 1620
QY 1621 GGGGCCCCCAGAGAGATTTGAAGGCGCTCAGTCAAGAGATGTGAGCATGTGAGCT 1680
DB 1621 GGGGCCCCCAGAGAGATTTGAAGGCGCTCAGTCAAGAGATGTGAGCATGTGAGCT 1680
QY 1681 AGCTTATGAGAGCTGGGAGGAGGATGCGGTGTGCGCTGAGTGTGGCTGAGAGTCAAGAA 1740
DB 1681 AGCTTATGAGAGCTGGGAGGAGGATGCGGTGTGCGCTGAGTGTGGCTGAGAGTCAAGAA 1740
QY 1741 CTTCGTGAGACCCACAGTGGAGGAGGCTGAAAGCGGAGAGAGCGGCAACCGCGTGGGAGC 1800
DB 1741 CTTCGTGAGACCCACAGTGGAGGAGGCTGAAAGCGGAGAGAGCGGCAACCGCGTGGGAGC 1800
QY 1801 TGGATCTTACGCGAGCGGAGAGGAGCTGAGACAGCTGGGGCTGGGGCTACAGGCGGC 1860
DB 1801 TGGATCTTACGCGAGCGGAGAGGAGCTGAGACAGCTGGGGCTGGGGCTACAGGCGGC 1860
QY 1861 ATCCCAAGGCTACCTGCTGAGTGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1920
DB 1861 ATCCCAAGGCTACCTGCTGAGTGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1920
QY 1921 TGGCTCTAGAGACCTGAGACCTGTTCTACGCTGTGGGAGCTGCTAGCTTCCACCTG 1980
DB 1921 TGGCTCTAGAGACCTGAGACCTGTTCTACGCTGTGGGAGCTGCTAGCTTCCACCTG 1980
QY 1981 GCGTGGGGCCCACTCTCTTGTGGCCCGAGCCCTGTGGGATATCCCGCTGGCCAGG 2040
DB 1981 GCGTGGGGCCCACTCTCTTGTGGCCCGAGCCCTGTGGGATATCCCGCTGGCCAGG 2040
QY 2041 AGCAGGACGAGGCTATCCCGTTCACCCCAAGAGAACTGCGCTCAGTAAGGAGGA 2100
DB 2041 AGCAGGACGAGGCTATCCCGTTCACCCCAAGAGAACTGCGCTCAGTAAGGAGGA 2100
QY 2101 TGCCCTCTGAGACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 2138
```

```

Db 2101 TGCCCCCTGCAGACAAAAA 2138
RESULT 2
ARL19934 2138 bp DNA PAT 16-MAY-2001
LOCUS ARL19934
DEFINITION Sequence 1 from patent US 6153430.
ACCESSION ARL19934
VERSION ARL19934.1 GI:14102633
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2138)
AUTHORS Pastan,I. and Chang,K.
TITLE Nucleic acid encoding mesothelin, a differentiation antigen present
on mesothelium, mesotheliomas and ovarian cancers
JOURNAL Patent: US 6153430-A 1 28-NOV-2000;
FEATURES
source 1. 2138
location/Qualifiers
BASE COUNT 385 a 722 c 661 g 370 t
ORIGIN
Query Match 100.0%; Score 2138; DB 6; Length 2138;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGGAATTCGGGTGGCCGGCCACTCCCTGCTGTGTGAGCGCCGGACAGAGAGCTAACCGT 60
Db 1 AGGAATTCGGGTGGCCGGCCACTCCCTGCTGTGTGAGCGCCGGACAGAGAGCTAACCGT 60
QY 61 GGACCCAGGAGGCTCCCTCCCTGGATCTACAGAGACATGCGCTTGCAAGCGCTGCAC 120
Db 61 GGACCCAGGAGGCTCCCTCCCTGGATCTACAGAGACATGCGCTTGCAAGCGCTGCAC 120
QY 121 CCCGTTGGTCTGTGGGGAGCGCCCTGCGAGCTCCTGCTGTGTGAGCGCTGGA 180
Db 121 CCCGTTGGTCTGTGGGGAGCGCCCTGCGAGCTCCTGCTGTGTGAGCGCTGGA 180
QY 121 CCCGTTGGTCTGTGGGGAGCGCCCTGCGAGCTCCTGCTGTGTGAGCGCTGGA 180
Db 121 CCCGTTGGTCTGTGGGGAGCGCCCTGCGAGCTCCTGCTGTGTGAGCGCTGGA 180
QY 181 TGGTGATCCCGCGAGACCTGCTGTGAGAGACAGGAGCGAGTGTGCCCTGGGG 240
Db 181 TGGTGATCCCGCGAGACCTGCTGTGAGAGAGAGGAGCGAGTGTGCCCTGGGG 240
QY 241 GGAATCTCGAACAACCCCAATTAATTTCCAGCTCTCCCTCCCAACTCTTGGCTTC 300
Db 241 GGAATCTCGAACAACCCCAATTAATTTCCAGCTCTCCCTCCCAACTCTTGGCTTC 300
QY 301 CCGTGTGCGAGGTGTCCGGCTGAGCAGGAGGCTGTCCGGAGCTGGCTGTGGCTTG 360
Db 301 CCGTGTGCGAGGTGTCCGGCTGAGCAGGAGGCTGTCCGGAGCTGGCTGTGGCTTG 360
QY 361 GCACAGAGAATGTCAAGCTCTCAACAGAGACGCTGCGCTGTGCGTCAACGGCTCTCT 420
Db 361 GCACAGAGAATGTCAAGCTCTCAACAGAGAGCTGCGCTGTGCGTCAACGGCTCTCT 420
QY 421 GAGCCCCCGAGAGACCTGAGAGCGCTCCCATTTGACCTGCTATTTCTCAACCCAGAT 480
Db 421 GAGCCCCCGAGAGACCTGAGAGCGCTCCCATTTGACCTGCTATTTCTCAACCCAGAT 480
QY 481 GCGTTCCTGCGGGCCCGAGGCTGTGACCGCTTCTTCTCCCGCATAGAGAGGCAATGTG 540
Db 481 GCGTTCCTGCGGGCCCGAGGCTGTGACCGCTTCTTCTCCCGCATAGAGAGGCAATGTG 540
QY 541 GACCTGCTCCCGAGGGGGCTCCGAGCGACAGCGGCTGTGCTGCGGCTCTGAGCTGC 600
Db 541 GACCTGCTCCCGAGGGGGCTCCGAGCGACAGCGGCTGTGCTGCGGCTCTGAGCTGC 600
QY 601 TGGGGTGTGCGGGGGTCTCTCTCTGAGCGAGGCTGATGTGCGGGCTCTGAGAGGCTTGGCT 660
Db 601 TGGGGTGTGCGGGGGTCTCTCTCTGAGCGAGGCTGATGTGCGGGCTCTGAGAGGCTTGGCT 660

```

```

QY 661 TGGACCTGCTGGGGGCTTTGTGGCCGAGTGGGCCGAGTGTCTACCCGGCTGCTG 720
Db 661 TGGACCTGCTGGGGGCTTTGTGGCCGAGTGGGCCGAGTGTCTACCCGGCTGCTG 720
QY 721 AGCTGCCCGGAGCCCTTGAGCAGGACCAAGCAGAGGAGCGGCGCTCTGAGAGGC 780
Db 721 AGCTGCCCGGAGCCCTTGAGCAGGACCAAGCAGAGGAGCGGCGCTCTGAGAGGC 780
QY 781 GGGGAGCCCCCTTACGGCCCCCGTCCAGATGTGTCTCTCAGAGTGTGAGCTGTGGG 840
Db 781 GGGGAGCCCCCTTACGGCCCCCGTCCAGATGTGTCTCTCAGAGTGTGAGCTGTGGG 840
QY 841 GGCTGTGCTGGCCGAGGAGCGCCATCTCCGACATCTCCGAGGATGCTGTGGC 900
Db 841 GGCTGTGCTGGCCGAGGAGCGCCATCTCCGACATCTCCGAGGATGCTGTGGC 900
QY 901 GCGTGGCGGCAAGCCTCTCTGAGGAGCCATCTGGCGGAGCGCTGAGAGCAATCTCTC 960
Db 901 GCGTGGCGGCAAGCCTCTCTGAGGAGCCATCTGGCGGAGCGCTGAGAGCAATCTCTC 960
QY 961 CGGCGCGGCTTCCGGCGGAGAGTGGAGAGACAGCTCTCTTACAGCAAGAGCCCGC 1020
Db 961 CGGCGCGGCTTCCGGCGGAGAGTGGAGAGACAGCTCTCTTACAGCAAGAGCCCGC 1020
QY 1021 GAGATAGACAGAGGCTGATCTTCTACAAAGTGGGAGCTGGAAGCTGTGGATGCG 1080
Db 1021 GAGATAGACAGAGGCTGATCTTCTACAAAGTGGGAGCTGGAAGCTGTGGATGCG 1080
QY 1081 GCGCTGTGCGCACCCAGATAGAGCGCGCTGTAAGCCATCTCCCTTACCTACAGAGAGCTG 1140
Db 1081 GCGCTGTGCGCACCCAGATAGAGCGCGCTGTAAGCCATCTCCCTTACCTACAGAGAGCTG 1140
QY 1141 GAGCTCTTAAAGCATTAACGTGATGAGCTCTACCCAGAGTTACCCGATCTGTATC 1200
Db 1141 GAGCTCTTAAAGCATTAACGTGATGAGCTCTACCCAGAGTTACCCGATCTGTATC 1200
QY 1201 CAGACAGTGGGCTACCTCTTCTCAAGATGAGCCCTGAGAGCATTCGAGAGTGAATG 1260
Db 1201 CAGACAGTGGGCTACCTCTTCTCAAGATGAGCCCTGAGAGCATTCGAGAGTGAATG 1260
QY 1261 ACGTCCCTGAGAGACCTGTAAGGCTTTGCTTGAAGTGCAGCAAGGCGACGAATGAGTCT 1320
Db 1261 ACGTCCCTGAGAGACCTGTAAGGCTTTGCTTGAAGTGCAGCAAGGCGACGAATGAGTCT 1320
QY 1321 CAGGCTCTGCGGGCGCCCTCCCAAGGTGGCCACCTGATGACAGCGTTTGTGAAGGGA 1380
Db 1321 CAGGCTCTGCGGGCGCCCTCCCAAGGTGGCCACCTGATGACAGCGTTTGTGAAGGGA 1380
QY 1381 AGGGGCGAGTAGACAAAGACACCTAGACACCTGAGCGGCTTCTACCTGTGGTACTCTG 1440
Db 1381 AGGGGCGAGTAGACAAAGACACCTAGACACCTGAGCGGCTTCTACCTGTGGTACTCTG 1440
QY 1441 TGTCTCTTACAGCCCGGAGGACTGAGCTCTGCTGCCCGCCAGCAGCATCTGGGCGTCA 1500
Db 1441 TGTCTCTTACAGCCCGGAGGACTGAGCTCTGCTGCCCGCCAGCAGCATCTGGGCGTCA 1500
QY 1501 CCCCAGAGACCTGTGACAGCTGTGACCCAAAGGAGCTGTGAGACGTCCTTACCCAAAGCCCGC 1560
Db 1501 CCCCAGAGACCTGTGACAGCTGTGACCCAAAGGAGCTGTGAGACGTCCTTACCCAAAGCCCGC 1560
QY 1561 CTTGCTTTCCAGAACATGAAACGGGTCGGAATACTTCTGGAAGATCAATCTTCTGTGGT 1620
Db 1561 CTTGCTTTCCAGAACATGAAACGGGTCGGAATACTTCTGGAAGATCAATCTTCTGTGGT 1620
QY 1621 GGGGCCCCCAGAGAGATTTGAAGGCGCTAGTCAGAGAAATGTGAGCATGACTTGGCC 1680
Db 1621 GGGGCCCCCAGAGAGATTTGAAGGCGCTAGTCAGAGAAATGTGAGCATGACTTGGCC 1680
QY 1681 ACGTTCAATGAAGCTGTGGAGGAGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
Db 1681 ACGTTCAATGAAGCTGTGGAGGAGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740

```

OY	1741	CTTCTGGGACCCCAAGCTGTAGAGGCGCTTGAAAGGGGAGGAGACGGCACCGCCTGTGGGCAC	1800
Db	1741	CTTCTGGGACCCCAAGCTGTAGAGGCGCTTGAAAGGGGAGGAGACGGCACCGCCTGTGGGCAC	1800
OY	1801	TGATCTCTAGCGGACAGCGGACAGCACTGTGAACACGCTGTGGGGCTACTACAGGGCGGC	1860
Db	1801	TGATCTCTAGCGGACAGCGGACAGCACTGTGAACACGCTGTGGGGCTACTACAGGGCGGC	1860
OY	1861	ATCCCCAACGGCTACCTGTGTCTTAAGCTTAGACCTTAGACCTGTGAAGACCCCTCTCGAGGACGCC	1920
Db	1861	ATCCCCAACGGCTACCTGTGTCTTAAGCTTAGACCTGTGAAGACCCCTCTCGAGGACGCC	1920
OY	1921	TGCCCTTCAGGACACCTGGACACTGTTCACACGCGCTGGACACTGTCTTAGACCCTCACCCCTG	1980
Db	1921	TGCCCTTCAGGACACCTGGACACTGTTCACACGCGCTGGACACTGTCTTAGACCCTCACCCCTG	1980
OY	1981	GCTTGAGGGCCCCCTACTCCCTTTGTCTGTGGCCCCAAGCCCTGTGTGGGATCCCCTGTGGCCAGG	2040
Db	1981	GCTTGAGGGCCCCCTACTCCCTTTGTCTGTGGCCCCAAGCCCTGTGTGGGATCCCCTGTGGCCAGG	2040
OY	2041	AGCAGGACAGGGGTATGTCCCGCTGCCACCCCAAAGAACAACCTGGCGCTCAGTAACAGGGGAACA	2100
Db	2041	AGCAGGACAGGGGTATGTCCCGCTGCCACCCCAAAGAACAACCTGGCGCTCAGTAACAGGGGAACA	2100
OY	2101	TGCCCCCTGCAGACAAAAAAGAAAAAAAAAAAAAAAAAAAAA 2138	
Db	2101	TGCCCCCTGCAGACAAAAAAGAAAAAAAAAAAAAAAAAAAAA 2138	
RESULT	3		
LOCUS	HSU40434	HSU40434 2114 bp mRNA PRI 19-JAN-1996	
DEFINITION		Human mesothelin or CAK1 antigen precursor mRNA, complete cds.	
ACCESSION		U40434	
VERSION		U40434.1 GI:1145723	
KEYWORDS			
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Graniteta; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLE		1 (bases 1 to 2114)	
JOURNAL		Chang, K. and Pastan, I.	
MEDLINE		Molecular cloning of mesothelin, a differentiation antigen present	
REFERENCE		Proc. Natl. Acad. Sci. U.S.A. 93 (1), 146-140. (1996).	
AUTHORS		Chang, K.	
TITLE		Direct Submmission	
JOURNAL		Submitted (09-NOV-1995) Kai Chang, Laboratory of Molecular Biology,	
MEDLINE		National Cancer Institute, Building 37, Room 4B19, 37 Convent	
REFERENCE		Drive, MSC4235, Bethesda, MD MD20892-4235, USA	
AUTHORS		Location/Ouilifiers	
TITLE		1. 2114	
FEATURES			
source		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="pcD3CAK1-9"	
		/chromosome="16"	
		/cell_line="KB, HeLa, OVCAR-3"	
		/cell_type="HeLa"	
		/tissue_type="ovarian cancers and squamous cell carcinoma."	
		100..1986	
CDS		/codon_start=1	
		/product="mesothelin or CAK1 antigen precursor"	
		/protein_id="AAC50346.1"	
		/db_xref="GI:1145724"	
		/translation="MALORLDPCWSCSGDRPSLFLFSLGWHPARTLAGETTESA	
		PGLGVLPPIHNISLSROLGFPCAEGASGSTRVEHLAVALKQKNVKSTEOLRCI	
		AHRILSPEDIDALPLDLFLFNDAFGSPGAPCRFSRTRKANVDLLPRCAPRORI	
		LPLALACWGNVGSLLSEADNRALGCLACDIDPGRVAASAENVLEPLRVSCPEPLDDQC	
		EAAMAAQGGPPPTGPSTVSVMADLRGLLPVLGDPITRSIDGGIVAAWRSSSR	
		PSMVAQRITIRPRPREVEVTACPSPGSKAREIDESLIIFYKWELNLCVDAALLATQM	

[illegible]





FT 3'UTR  
FT misc-feature  
FT  
FT  
Location/Qualifiers  
1. 2129  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 389 a 716 c 657 g 367 t  
ORIGIN

Query Match 93.3%; Score 1994; DB 6; Length 2129;  
Best Local Similarity 97.0%; Pred. No. 0;  
Matches 2075; Conservative 0; Mismatches 35; Indels 30; Gaps 3;

potentiating  
activity  
1966..2129  
1873  
/note="The nucleotide at position 1873 of cDNA  
in other  
clone pKPO21 is A"

5 ATTCGGTGGCGGCACTCCGTCGTCTGTGACGGCGGACAGAGAGCTACGGTGGAC 64  
2 AATTGGGACAGGACCACTCCGTCGTCTGTGACGGCGGACAGAGAGCTACGGTGGAC 61  
65 CCAGGGGACCTCCCTCCCTGGGATCTACAGAGACCATGGCTTG-CAAGGCTGACCCC 123  
62 CCAGGTCCTCCCTCCCTGGGATCTACAGAGACCATGGCTTGCAAGGCTGACCCC 121  
124 TGTGTGTCCTGTGGGACCGCC-----TGAGAGCTCTCTCTGCTGCTTTCAGCTG 178  
122 TGTGGGGTCTGTGGGACCGCCCGCCCTGGGACGCTCTCTGCTGCTTTCAGCTG 181  
179 GATGGGTGATCCCGGACGACCTGTGGAGAGACAGGAGGAGTCTGCCCGCTGG 238  
182 GATGGGTGACACCTCCGACGACCTGTGGAGAGACAGGAGGAGGAGTCTGCCCGCTGG 241  
239 GGGAGTCTGACACCCCGCATACATTTCCAGCTCTCCCTCCGCACTCTGCTGCT 298  
242 ACGGAGTCTGACACCCCGCATACATTTCCAGCTCTCCCTCCGCACTCTGCTGCT 301  
299 TCCGCTGTGGGAGTGTCCGCTGTGACAGAGAGGCTGCTGGGAGCTGTGGCT 358  
302 TCCGCTGTGGGAGTGTCCGCTGTGACAGAGAGGCTGCTGGGAGCTGTGGCT 361  
359 TGGCAGAGAGATGTCAAGCTCTCAACAGAGAGTGTGGCTGTGGCTGTGGCTCT 418  
362 TGGCAGAGAGATGTCAAGCTCTCAACAGAGAGTGTGGCTGTGGCTGTGGCTCT 421  
419 CTGAGCCCCCGAGAGAGCTGTGAGGCGCTCCCATTTGAGAGCTGTCTATCTCAACCAG 478  
422 CTGAGCCCCCGAGAGAGCTGTGAGGCGCTCCCATTTGAGAGCTGTCTATCTCAACCAG 481  
479 ATGGCTGTGGGAGGCGCCAGGCTGTGACAGAGGCTTCTCTCCGCTATGACAGAGGCAATG 538  
482 ATGGCTGTGGGAGGCGCCAGGCTGTGACAGAGGCTTCTCTCCGCTATGACAGAGGCAATG 541  
539 TGGCAGAGAGATGTCAAGCTCTCAACAGAGAGTGTGGCTGTGGCTGTGGCTCT 598  
542 TGGCAGAGAGATGTCAAGCTCTCAACAGAGAGTGTGGCTGTGGCTGTGGCTCT 601  
599 GCTGGGGTGTGGGAGGCTGTCTGTGAGCAGAGGCTGATGTGGGGCTGTGGAGAGGCTGG 658  
602 GCTGGGGTGTGGGAGGCTGTCTGTGAGCAGAGGCTGATGTGGGGCTGTGGAGAGGCTGG 661  
659 CTTCGAGAGCTGTGGGAGGCTGTGTGGCGAGTGTGGCGGAAGTGTCTGACCCGGGCTGG 718  
662 CTTCGAGAGCTGTGGGAGGCTGTGTGGCGAGTGTGGCGGAAGTGTCTGACCCGGGCTGG 721  
719 TGAGCTCCCGGAGAGCCCTGAGACAGAGACGAGAGAGGAGCCAGGCTGTGGAGG 778  
722 TGAGCTCCCGGAGAGCCCTGAGACAGAGACGAGAGAGGAGCCAGGCTGTGGAGG 781  
779 GCGGGAGAGCCCTGAGAGGCGCCGTCGACATGTGTGTCTGTCACAGATGAGGCTGTGC 838

Db 782 GCGGGAGAGCCCGCTACGCGCCCGCTGCGACATGTGTCTCCACGATGAGCGCTTGC 841  
Qy 839 GGGGCTGCTCCCGTGTGGGACAGCCCATATCCGAGCATCCGAGGAGCATGCGG 898  
Db 842 GGGGCTGCTCCCGTGTGGGACAGCCCATATCCGAGCATCCGAGGAGCATGCGG 901  
Qy 899 CCGGCTGCGGCAACGCTCTCTGGGAGCCCATCTTGGGCGGAGCTGMAAGGACATCC 958  
Db 902 CCGGCTGCGGCAACGCTCTCTGGGAGCCCATCTTGGGCGGAGCTGMAAGGACATCC 961  
Qy 956 TCCGGCGCGGCTTCCGGGAGAGTGGAGAGACAGCTGTCTTCAAGGCAAGAGGCC 1018  
Db 962 TCCGGCGCGGCTTCCGGGAGAGTGGAGAGACAGCTGTCTTCAAGGCAAGAGGCC 1021  
Qy 1019 GCGAGATGAGAGAGCTCTCTTCTACAGAGTGGAGCTGGAAGCCGTGCTGATG 1078  
Db 1022 GCGAGATGAGAGAGCTCTCTTCTTCAAGAGTGGAGCTGGAAGCCGTGCTGATG 1081  
Qy 1079 GGGCTCTGCTGGCACCAGATGGACCGGCTGAAAGCCATCCCTTACCTAGAGAGC 1138  
Db 1082 GGGCTCTGCTGGCACCAGATGGACCGGCTGAAAGCCATCCCTTACCTAGAGAGC 1141  
Qy 1139 TGGAGTCTTAAGCATTAAGTATGATGATGATCTTACCCAGAGGTTACCCGAGTCTGA 1198  
Db 1142 TGGAGTCTTAAGCATTAAGTATGATGATGATCTTACCCAGAGGTTACCCGAGTCTGA 1201  
Qy 1199 TCCAGACACTGGGCTACTCTTCTCAAGATGAGCCCTGAGAGATTCGCAAGTGAATG 1258  
Db 1202 TCCAGACACTGGGCTACTCTTCTCAAGATGAGCCCTGAGAGATTCGCAAGTGAATG 1261  
Qy 1259 TGAAGTCTTGAAGCCTGGAAGCTTCTTGAAGTGAAGGAGGAGCAAGATGATG 1318  
Db 1262 TGAAGTCTTGAAGCCTGGAAGCTTCTTGAAGTGAAGGAGGAGCAAGATGATG 1321  
Qy 1319 CTGAGGCTCTGCGGCGCCCTCCACAGCTGGCCACCTGATGAGCCGTTTGTGAAG 1378  
Db 1322 CT-----CAGGTGGCCACCTGATGAGCCGTTTGTGAAG 1387  
Qy 1379 GAAGGGGACAGTACAGAAAGACACCTTACACCTGTGACCGCTTACCTGGTATC 1438  
Db 1388 GAAGGGGACAGTACAGAAAGACACCTTACACCTGTGACCGCTTACCTGGTATC 1447  
Qy 1439 TGTGCTCTCAGCCCGGAGAGCTGAGCTTCTGTCGCCCGCCAGCAGATCTGGCGGTCA 1498  
Db 1448 TGTGCTCTCAGCCCGGAGAGCTGAGCTTCTGTCGCCCGCCAGCAGATCTGGCGGTCA 1477  
Qy 1499 GGGCCAGAGAGCTGGAAGAGTGAAGCCAGAGAGAGTGTGAGCTCTATCCAGAGGCC 1558  
Db 1478 GGGCCAGAGAGCTGGAAGAGTGAAGCCAGAGAGAGTGTGAGCTCTATCCAGAGGCC 1537  
Qy 1559 GCTTGTCTTCCAGAGATGAAAGGGTCCAGATCTTCTGAAAGATCCAGTCTTCTGG 1618  
Db 1538 GCTTGTCTTCCAGAGATGAAAGGGTCCAGATCTTCTGAAAGATCCAGTCTTCTGG 1597  
Qy 1619 GTGGGGCCCCAGGAGATTTGAAGGCTCACTCAGCAGATGTAAGCATGAGCTTGG 1678  
Db 1598 GTGGGGCCCCAGGAGATTTGAAGGCTCACTCAGCAGATGTAAGCATGAGCTTGG 1657  
Qy 1679 CCAAGTCTATGAAGCTCGGAGAGATGCGGTGCTGCGCTTATCTGTGAGTGTGAGA 1738  
Db 1658 CCAAGTCTATGAAGCTCGGAGAGATGCGGTGCTGCGCTTATCTGTGAGTGTGAGA 1717  
Qy 1739 AACTTGTGGAGCCCGAGAGTGTGAAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGG 1798  
Db 1718 AACTTGTGGAGCCCGAGAGTGTGAAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGG 1777  
Qy 1799 ACTGATCTAAGGAGGAGGAGAGCAGCTGAGACAGCTTGGGCTGGGCTTACAGGGCG 1858  
Db 1778 ACTGATCTAAGGAGGAGGAGGAGGAGCAGCTGAGACAGCTTGGGCTGGGCTTACAGGGCG 1837  
Qy 1859 GCATCCCAAGGCTACTGTGTCTAGACCTGACGCTGACAGAGACCTCTCGGGAGAGC 1918  
Db 1838 GCATCCCAAGGCTACTGTGTCTAGACCTGACGCTGACAGAGAGGAGGAGGAGGAGGAGC 1897

Query Match	Best Local Similarity	Matches 2075;	Conservative	0;	Mismatches	35;	Indels	30;	Gaps
93.3%; Score 1994;	97.0%; Pred. No. 0;								
DB 6;	Length 2129;								
5	ATTCCGCTGCGCCGACCTCCCGCTGCTGTGAGCGCGGACAGAGAGCTACCGGTGAC	64							
DB	2 AATTGCGGACGAGGCGCACCTCCGCTGCTGTGAGCGCGGACAGAGAGCTACCGGTGAC	61							
OY	65 CCAGGTCGCTCCCGCTCCCGGTATACACAGACATGGCCCTTG-CAAGGCTCGACCCC	123							
DB	62 CCAGGTCGCTCCCGCTCCCGGTATACACAGACATGGCCCTTGCAAGGCTCGACCCC	121							
OY	124 TGTGTCCTGTTGGGAGGACGCGC-----TGGAGCCTCCTGTTCCGCTCTTCAGCCTCG	178							
DB	122 TGTGTTGGGTCGTGTGGAGACCCCGGCTCGGAGCCTCCTGTTCCGCTCTTCAGCCTCG	181							
OY	179 GATGGGTGATCCCGCGAGAGACCTGCTGAGAGAGAGAGAGATGTCGCCCTCGG	238							
DB	182 GATGGGTGATCCCGCGAGAGACCTGCTGAGAGAGAGAGAGAGATGTCGCCCTCGG	241							
OY	239 GGGGAGTCTGACACCCCGCATATCTTCAGGCTCTCCCTCGGCAACCTCTTGCT	298							
DB	242 ACGGAGTCTGACACCCCGCATATCTTCAGGCTCTCCCTCGGCAACCTCTTGCT	301							
OY	239 TCCCGTGTGCGAGAGTGTCCGGCTGAGACAGGAGAGCTGTCCGGAGAGTGTGTGAGCT	358							
DB	302 TCCCGTGTGCGAGAGTGTCCGGCTGAGACAGGAGAGCTGTCCGGAGAGTGTGTGAGCT	361							
OY	359 TGGCAGACAGAGATGTCAAGCTCTCAACAGAGACAGCTGGCTGTGCTGTCAACGGCTCT	418							
DB	362 TGGCAGACAGAGATGTCAAGCTCTCAACAGAGACAGCTGGCTGTGCTGTCAACGGCTCT	421							
OY	419 CTGAGCCCCCGGAGACTGTGAGAGCCCTCCATTGTGAGCTGTGCTATTCTCAACCCAG	478							

D	b	422	CTGAGCCCCCGAGAGACTGGAGGCCCTCCATTGGACCGGTGCTATTTCTCAACCCAG	481
Q	y	479	ATGCGTTCTCGGGCCCCAGGCGCTGCACCCGTTTCTTCTCCGCAATCAGAAAGCCAAATG	538
D	b	482	ATGGGTCTCTCGGGGCCCCAGGCGCTGCACCCGTTTCTTCTCCGCAATCAGAAAGCCAAATG	541
Q	y	539	TGGACGTGCTCCGAGGGGGGGCTCCGAGAGAGAGGGGTGGTCCCTGTGGGCTCTGGGCT	598
D	b	542	TGGACGTGCTCCGAGGGGGGGCTCCGAGAGAGAGGGGTGGTCCCTGTGGGCTCTGGGCT	601
Q	y	599	GCTGGGAGTGTCCGGGGGTCTCTGCTGAGCAGAGCTCATGTGGGGGTCTGGGAGGCTGG	658
D	b	602	GCTGGGAGTGTCCGGGGGTCTCTGCTGAGCAGAGCTCATGTGGGGGTCTGGGAGGCTGG	661
Q	y	659	CTTGGACCTGCTCGGGCGCTTTGTGGCCGAGTGGCCGGAAGTGTCTTACCCCGGCTGG	718
D	b	662	CTTGGACCTGCTCGGGCGCTTTGTGGCCGAGTGGCCGGAAGTGTCTTACCCCGGCTGG	721
Q	y	719	TGAGCTGCCCCGGAGCCCTTGAGACAGAGCAGAGAGGAGCAGAGGGGCTCTGACAG	778
D	b	722	TGAGCTGCCCCGGAGCCCTTGAGACAGAGCAGAGAGGAGCAGAGGGGCTCTGACAG	781
Q	y	779	GCGGGGAGACCCCTTACAGGCCCCCGCTGCACATAGTGTCTTCTTCCAGATGAGACGCTTGC	838
D	b	782	GCGGGGAGACCCCTTACAGGCCCCCGCTGCACATAGTGTCTTCCAGATGAGACGCTTGC	841
Q	y	839	GAGGCTGTGCTCCCGCTGTGGGCCAGCCCATCATCCGACAGATCCCGCAGGGCATCTGG	898
D	b	842	GAGGCTGTGCTCCCGCTGTGGGCCAGCCCATCATCCGACAGATCCCGCAGGGCATCTGG	901
Q	y	899	CCGGTGGCGGCAACGCTCTTGGGACCATCTGCGGGGCGAGCTTGAAGGACCATCC	958
D	b	902	CCGGTGGCGGCAACGCTCTTGGGACCATCTGCGGGGCGAGCTTGAAGGACCATCC	961
Q	y	959	TCCGGCCCGGGTTCCGCGGGGAGGTGAGAGAGACAGCTGTCTTCTTACGGCAAGAGGCC	1018
D	b	962	TCCGGCCCGGGTTCCGCGGGGAGGTGAGAGAGACAGCTGTCTTCTTACGGCAAGAGGCC	1021
Q	y	1019	GCGAGATAGACAGAGACCTCATCTTCTACAAAGATGGGAGCTGGAGGCTCTCGTGATG	1078
D	b	1022	GCGAGATAGACAGAGACCTCATCTTCTACAAAGATGGGAGCTGGAGGCTCTCGTGATG	1081
Q	y	1079	GCGGCTCTGTGGGACCCAGATGGAGACGCGATGAAGCCCATCCCTTCACTACGAGAGAC	1138
D	b	1082	GCGGCTCTGTGGGACCCAGATGGAGACGCGATGAAGCCCATCCCTTCACTACGAGAGAC	1141
Q	y	1139	TGGAGCTCTTAAAGCATTAACTGGATAGCTCTTACCCACAGAAGTGTACCCCGAGTGTGCA	1198
D	b	1142	TGGAGCTCTTAAAGCATTAACTGGATAGCTCTTACCCACAGAAGTGTACCCCGAGTGTGCA	1201
Q	y	1199	TCCAGACCTGTGGGCTACCTCTTCTCAAGATGAGCCCTTGAGGACATTTGCCAAGTGAATG	1258
D	b	1202	TCCAGACCTGTGGGCTACCTCTTCTCAAGATGAGCCCTTGAGGACATTTGCCAAGTGAATG	1261
Q	y	1259	TGAGCTCCCTGGAGACCTCTAAGGCTTTGTGAACCTCAAAAGGGGACGAAATGAGTC	1318
D	b	1262	TGAGCTCCCTGGAGACCTCTAAGGCTTTGTGAACCTCAAAAGGGGACGAAATGAGTC	1321
Q	y	1319	CTCAGAGCTCTCGCGGCGCCCTCCCAATAGGTGGCCACCCCTGATCGACCCCTTTGTGAAG	1378
D	b	1322	CTCAGAGCTCTCGCGGCGCCCTCCCAATAGGTGGCCACCCCTGATCGACCCCTTTGTGAAG	1381
Q	y	1379	GAAAGGGGACCTAGACAAAGACACCCCTAGACACCCCTGACCGGCTTCTTACCCCTGGGTAC	1438
D	b	1388	GAAAGGGGACCTAGACAAAGACACCCCTAGACACCCCTGACCGGCTTCTTACCCCTGGGTAC	1441
Q	y	1439	TGTGCTCCTTAGCCCCGAGAGAGCTGAGCTTCGTGCCCCCAGACAGATCTGGGCGGTCA	1498
D	b	1448	TGTGCTCCTTAGCCCCGAGAGAGCTGAGCTTCGTGCCCCCAGACAGATCTGGGCGGTCA	1501
Q	y	1499	GAGCCGAGGAGCTGGACAGAGTGAACCAAGGAGAGTGGAGCGTCTTATCCCAAGGGCC	1558
D	b	1478	GAGCCGAGGAGCTGGACAGAGTGAACCAAGGAGAGTGGAGCGTCTTATCCCAAGGGCC	1561

QY 1559 GCCTTCTTCCAGAACATGACGCGGTCCGAAATCTGTCGTAAGATCCAGTCTCTTG 1618  
 Db 1538 GCCTTCTTCCAGAACATGACGCGGTCCGAAATCTGTCGTAAGATCCAGTCTCTTG 1597  
 QY 1619 GTGGGGCCCCCAGCAGAGATTTGAAGCGCTCAGTCACAGATGTGAGCATGCGTGG 1678  
 Db 1598 GTGGGGCCCCCAGCAGAGATTTGAAGCGCTCAGTCACAGATGTGAGCATGCGTGG 1657  
 QY 1679 CCAGCTCATGAGCTGGGAGCGAGTGGGTGGTGGCTGGCTGAGTGGCTGGAGTGGAGA 1738  
 Db 1658 CCAGCTCATGAGCTGGGAGCGAGTGGGTGGTGGCTGGCTGAGTGGCTGGAGTGGAGA 1717  
 QY 1739 AACTTCTTGGAGCCCAACGTCGAGGCGCTGAAAGCGGAGAGCGGAGCGCGCGGCTGGCGG 1798  
 Db 1718 AACTTCTTGGAGCCCAACGTCGAGGCGCTGAAAGCGGAGAGCGGAGCGCGCGGCTGGCGG 1777  
 QY 1799 ACTGATCTTACGCGACGCGGACGACGCTGAGACGCTGGGCTGGGCTGACAGGCGG 1858  
 Db 1778 ACTGATCTTACGCGACGCGGACGACGCTGAGACGCTGGGCTGGGCTGACAGGCGG 1837  
 QY 1859 GCATCCCCCAAGGGTACTGGTCTGAGACCTCAGGCTGAGAGAGAGAGAGAGAGAGAGAG 1918  
 Db 1838 GCATCCCCCAAGGGTACTGGTCTGAGACCTCAGGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1897  
 QY 1919 CCTGCTCTTCTGAGACCTGAGCTGCTTCTCAGCCGCTGAGCTGCTTCTGAGCTGAGCTGAG 1978  
 Db 1898 CCTGCTCTTCTGAGACCTGAGCTGCTTCTCAGCCGCTGAGCTGCTTCTGAGCTGAGCTGAG 1957  
 QY 1979 TGGCTGAGGCGCCCACTCTGCTGCTGCGCCGACGCTGCTGCGGAGTCCCGCTGCGCA 2038  
 Db 1958 TGGCTGAGGCGCCCACTCTGCTGCTGCGCCGACGCTGCTGCGGAGTCCCGCTGCGCA 2017  
 QY 2039 GGAGAGGACGAGGCGGATCCCGTTCACGCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2098  
 Db 2018 GGAGAGGACGAGGCGGATCCCGTTCACGCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2077  
 QY 2099 CATGCCCCCTGACAGCAAAAAAAAAAAAAAAAAAAAAA 2138  
 Db 2078 CATGCCCCCTGACAGCAAAAAAAAAAAAAAAAAAAAAA 2117

RESULT 6  
 BC003512 LOCUS 2433 bp mRNA PRI 12-JUL-2001  
 DEFINITION Homo sapiens, mesothelin, clone MGC:10686 IMAGE:3611296, mRNA, complete cds.  
 ACCESSION BC003512  
 VERSION BC003512.1 GI:13097581  
 KEYWORDS MGC.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Srausberg, R.  
 Direct Submission  
 Submitted (20-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 Info@cgsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
 Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo

Sen Lee, Victor Ling, Carrie Mathewson, Candice McLaavy, Steven  
 Ness, Pawan Prabhu, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline  
 Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,  
 Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Zetter.  
 Clon distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRAL Plate: 13 Row: a Column: 21  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 1129078.  
 FEATURES  
 source  
 1. 2433  
 /organism="Homo sapiens"  
 /db\_xref="locusID:10232"  
 /db\_xref="taxon:9606"  
 /clone="MGC:10686 IMAGE:3611296"  
 /tissue\_type="Pancrreas; adenocarcinoma"  
 /clone\_id="NIH\_MGC\_39"  
 /lab\_host="DH10B-R"  
 /note="Vector: p0187"  
 418..2283  
 /codon\_start=1  
 /product="mesothelin"  
 /protein\_id="AAH03512.1"  
 /db\_xref="GI:13097582"  
 /translation="MALPRTAPRLIGSGTPTALSLFLFLSLGVOPRTLAETGGA  
 APDGVLANPNISSLSPLRLGPRCAVEGSLSTERRRELAVALAOKNVLSTBOLRC  
 LAHRLSEPPDLPLDLPLPDLFSPQKTRFSNITRANVLDLRLGRLPEROR  
 LEPALACWGVRSLSLEADVRLAGLACDLPFVAESAVLLPRLVSPGLDDO  
 OEARALALOGGPPPPSTVSYSTMALGLPLVIGLPIRSIPQIVAMROSSR  
 DPSWROPERTILRPREFREVEKTACPSGKAREIDSLIYKKMLEACVDALLATO  
 MDNRALNPPFYEDLVKHLDELIVPGYSESVIOLGYFLKMSPEDIRKWNLSLE  
 TKRALNPNKMSPOVATLIDRPVGRDLDKDTLTATPPTGYSLSPEELSS  
 VPSSIRAVVPDRLDTCPRDLDVLRKALALONNNGSEYFKIOSFLGATPEDIK  
 ALSQONVSNLDFETMKRTDAVLPVVAEYQKLGLPVEELKAEERRRPVDMILQOR  
 QDDLDLGLDGLGGLPNCYLVLDLSVDEALSGTPCLGLPGLVPLVIALIASYLA"  
 BASE COUNT 431 a 823 c 755 g 424 t  
 ORIGIN  
 Query Match 92.7% Score 1981.8; DB 9; Length 2433;  
 Best Local Similarity 97.2% Pred. No. 0;  
 Matches: 2071; Conservative 0; Mismatches 27; Indels 33; Gaps 4;  
 QY 12 TGGCGGCGCATCCCGTCTGCTGAGCGCGGACAGAGACTCGGTGGAGCCACAGGT 71  
 Db 330 TGGCGGCGCATCCCGTCTGCTGAGCGCGGACAGAGACTCGGTGGAGCCACAGGT 389  
 QY 72 GCTTCCTCCCTGGAGATCTACAGAGACCATGGCTTG-CAAGGCTGACCCCTGTTGCT 130  
 Db 390 GCTTCCTCCCTGGAGATCTACAGAGACCATGGCTTGCAAGGCTGACCCCTGTTGCT 449  
 QY 131 CCTTGGGAGCCGCC-----TGGCAGCTTCCTGCTCTTCAAGCTCGGATGGGT 185  
 Db 450 GTCTGTGGGAGCCCGCCGCTTCCTGCTCTTCAAGCTCGGATGGGT 509  
 QY 186 GCATCCCGCGGAGACCGTGGCTGAGAGACAGGAGAGTCCGCCCGGGGGAGGT 245  
 Db 510 GCATCCCGCGGAGACCGTGGCTGAGAGACAGGAGTCCGCCCGGGGGAGGT 566  
 QY 246 CCTGACACCCCGCCATTAATTTCCAGCTTCCCTCGCCCACTCTTGGCTTCCCGGTG 305  
 Db 567 CCTGACACCCCGCCATTAATTTCCAGCTTCCCTCGCCCACTCTTGGCTTCCCGGTG 626  
 QY 306 TGGGAGGTGTCGGCTGAGACGAGAGCTGTCGGAGAGCTGGCTTGGCTTGGCACA 365  
 Db 627 TGGGAGGTGTCGGCTGAGACGAGAGCTGTCGGAGAGCTGGCTTGGCTTGGCACA 686  
 QY 366 GAAGAATGTCAAGTCTCAACAGAGAGAGCTGCTGCTGCTACCGGCTCTTGAGCC 425  
 Db 687 GAAGAATGTCAAGTCTCAACAGAGAGAGCTGCTGCTGCTACCGGCTCTTGAGCC 746

QY 426 CCCCCAGACCTGAGACGCCCTCCATTGGACCTGCTGCTATTCTTCAACCCAGATGCGTT 485  
 |||||||  
 Db 747 CCCCCAGACCTGAGACGCCCTCCATTGGACCTGCTGCTATTCTTCAACCCAGATGCGTT 806  
 |||||||  
 QY 486 CTTGGGGGCCAGGAGCTGACACCCCTTTCTCTCCGATCAGAGAGCCAAATGTGACCT 545  
 |||||||  
 Db 807 CTTGGGGGCCAGGAGCTGACACCCCTTTCTCTCCGATCAGAGAGCCAAATGTGACCT 866  
 |||||||  
 QY 546 GCTCCCAAGGGGGGGCTCCCGAGCAGACCGGCTGCTGCTGGGCTCTGGGCTCTGGG 605  
 |||||||  
 Db 867 GCTCCCAAGGGGGGGCTCCCGAGCAGACCGGCTGCTGCTGGGCTCTGGGCTCTGGG 926  
 |||||||  
 QY 606 TGTCTGGGGGGCTCTGCTGAGACGAGCTGATGTCGGGCTCTGGGAGCCCTGGCTTGGGA 665  
 |||||||  
 Db 927 TGTCTGGGGGGCTCTGCTGAGACGAGCTGATGTCGGGCTCTGGGAGCCCTGGCTTGGGA 986  
 |||||||  
 QY 666 CTTGCTTGGGGGGCTTTTGTGGCCGAGTGGCCGAAAGTCTGCTACCCGGCTGTGTGACCTG 725  
 |||||||  
 Db 987 CTTGCTTGGGGGGCTTTTGTGGCCGAGTGGCCGAAAGTCTGCTACCCGGCTGTGTGACCTG 1046  
 |||||||  
 QY 726 CCGGGGACCCCTGGACAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 785  
 |||||||  
 Db 1047 CCGGGGACCCCTGGACAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1106  
 |||||||  
 QY 786 ACCCCCTACAGGCCCCCGCTGACATGCTGTCTCCAGATGAGACGCTTGGGGGGCT 845  
 |||||||  
 Db 1107 ACCCCCTACAGGCCCCCGCTGACATGCTGTCTCCAGATGAGACGCTTGGGGGGCT 1166  
 |||||||  
 QY 846 GCTCCCGCTGTGGGGCCAGCCCATATCTCCGAGCATCCCGAGGAGCATCTGGCCGCTG 905  
 |||||||  
 Db 1167 GCTCCCGCTGTGGGGCCAGCCCATATCTCCGAGCATCCCGAGGAGCATCTGGCCGCTG 1226  
 |||||||  
 QY 906 GCGGCAACGCTCTCTGGGAGCCCATCTGGGAGGAGCTGAGGAGCATCTCCGCGC 965  
 |||||||  
 Db 1227 GCGGCAACGCTCTCTGGGAGCCCATCTGGGAGGAGCTGAGGAGCATCTCCGCGC 1286  
 |||||||  
 QY 966 GCGGCTCCGGGGAGGTGAGAGACAGCCTGTCTTCAAGCAAGAGGCGCGGAGAT 1025  
 |||||||  
 Db 1287 GCGGCTCCGGGGAGGTGAGAGACAGCCTGTCTTCAAGCAAGAGGCGCGGAGAT 1346  
 |||||||  
 QY 1026 AGACGAGAGCTCATCTTCTACAGAGAGTGGAGCTGGAACCTGCTGTGATCGGCGCT 1085  
 |||||||  
 Db 1347 AGACGAGAGCTCATCTTCTACAGAGAGTGGAGCTGGAACCTGCTGTGATCGGCGCT 1406  
 |||||||  
 QY 1086 GCTGGCCACCCAGATGAGACCGCGTGAAGCGCCATCCCTTCACTACAGAGAGTGAAGCT 1145  
 |||||||  
 Db 1407 GCTGGCCACCCAGATGAGACCGCGTGAAGCGCCATCCCTTCACTACAGAGAGTGAAGCT 1466  
 |||||||  
 QY 1146 CCTAAGCATTAAGTGAATGAGCTTACCCACAAAGTTACCCCGAGTGTGATCCAGCA 1205  
 |||||||  
 Db 1467 CCTAAGCATTAAGTGAATGAGCTTACCCACAAAGTTACCCCGAGTGTGATCCAGCA 1526  
 |||||||  
 QY 1206 CTTGGGCTACTCTTCTTCTCAAGATGAGCCCTGAGAGCATTTGGCAAGTGAATGTGACCTC 1265  
 |||||||  
 Db 1527 CTTGGGCTACTCTTCTTCTCAAGATGAGCCCTGAGAGCATTTGGCAAGTGAATGTGACCTC 1586  
 |||||||  
 QY 1266 CTTGGGAGACCTTGAAGGCTTGTGTTGAAGTGAAGAGGAGCAAGAAATGATCTCAGGC 1325  
 |||||||  
 Db 1587 CTTGGGAGACCTTGAAGGCTTGTGTTGAAGTGAAGAGGAGCAAGAAATGATCTCAGGC 1641  
 |||||||  
 QY 1326 TCTCTGGGGGGCTCTCCACAGTGGCGCACCTGATCGACGCTTTGTGAAGGAGAGGG 1385  
 |||||||  
 Db 1642 TCTCTGGGGGGCTCTCCACAGTGGCGCACCTGATCGACGCTTTGTGAAGGAGAGGG 1682  
 |||||||  
 QY 1386 CCAGCTAGACAAAGACACCTTACAGACCTTACCGCTTGTACCTGGGTACCTGTGCTC 1445  
 |||||||  
 Db 1683 CCAGCTAGACAAAGACACCTTACAGACCTTACCGCTTGTACCTGGGTACCTGTGCTC 1742  
 |||||||  
 QY 1446 CCTGAGCCCGGAGGAGCTGAGCTCGTCCGCCACAGAGCATCTGGGGGGGTCAGGCCCA 1505  
 |||||||  
 Db 1743 CCTGAGCCCGGAGGAGCTGAGCTCGTCCGCCACAGAGCATCTGGGGGGGTCAGGCCCA 1802  
 |||||||  
 QY 1506 GGACCTGAGACGTTGTGACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1565  
 |||||||

Db 1803 GGACCTGAGACAGTGTGACCCAGAGGAGCTGAGAGCTCTCTATATCCAAAGGCGCGGCTTGC 1862  
 |||||||  
 QY 1566 TTTCAGACATGAGAGGGGTCCGAATACTTGTGAGATCCAGTCTTCTTGGTGGGGC 1625  
 |||||||  
 Db 1863 TTTCAGACATGAGAGGGGTCCGAATACTTGTGAGATCCAGTCTTCTTGGTGGGGC 1922  
 |||||||  
 QY 1626 CCCCAGGAGATTGAAAGGGGCTCAGTACAGATGTGAGCATGTGACTGGGACGTT 1685  
 |||||||  
 Db 1923 CCCCAGGAGATTGAAAGGGGCTCAGTACAGATGTGAGCATGTGACTGGGACGTT 1982  
 |||||||  
 QY 1686 CATGAGCTGGGAGCGAGTGGGCTGCTCCCTTGAATGTGCTGAGTGCAGAAATCTT 1745  
 |||||||  
 Db 1983 CATGAGCTGGGAGCGAGTGGGCTGCTCCCTTGAATGTGCTGAGTGCAGAAATCTT 2042  
 |||||||  
 QY 1746 GGGACCCACATGAGAGGCTCTGAAGGCGGAGAGCGGACCGCCGGTGGGACTGGAT 1805  
 |||||||  
 Db 2043 GGGACCCACATGAGAGGCTCTGAAGGCGGAGAGCGGACCGCCGGTGGGACTGGAT 2102  
 |||||||  
 QY 1806 CCTAGGCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1865  
 |||||||  
 Db 2103 CCTAGGCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2162  
 |||||||  
 QY 1866 CAAGGCTACCTGCTCTAGACCTCAGGCTCAAGAGACCTCTGCGGGAGCGCCCTGCT 1935  
 |||||||  
 Db 2163 CAAGGCTACCTGCTCTAGACCTCAGGCTCAAGAGACCTCTGCGGGAGCGCCCTGCT 2222  
 |||||||  
 QY 1926 CCTAGAGACCTGAGACCTTCTTCCACCTGCTTGGGACCTGCTCTAGGCTTCCAGCTGCTG 1985  
 |||||||  
 Db 2223 CCTAGAGACCTGAGACCTTCTTCCACCTGCTTGGGACCTGCTCTAGGCTTCCAGCTGCTG 2282  
 |||||||  
 QY 1986 AGGGCCACCTCCCTTCTGCTGAGCCCTCAGGCTGTGGGAGATCCCGCTGCGGAGGAGGAG 2045  
 |||||||  
 Db 2283 AGGGCCACCTCCCTTCTGCTGAGCCCTCAGGCTGTGGGAGATCCCGCTGCGGAGGAGGAG 2342  
 |||||||  
 QY 2046 GCAGGGTGAATCCCGCTTCCACCCAGAGAACTCGGCTCAGTAAACGGGAACTGCC 2105  
 |||||||  
 Db 2343 GCAGGGTGAATCCCGCTTCCACCCAGAGAACTCGGCTCAGTAAACGGGAACTGCC 2402  
 |||||||  
 QY 2106 CTTGAGACAAAAA 2136  
 |||||||  
 Db 2403 CTTGAGACAAAAA 2433  
 |||||||

RESULT 7  
 HUMMPF 2093 bp mRNA PRI 01-FEB-2000  
 LOCUS Human mRNA for pre-pro-megakaryocyte potentiating factor, complete cds.  
 DEFINITION D49441  
 ACCESSION D49441 GI:1129078  
 VERSION 1  
 KEYWORDS MRF; megakaryocyte potentiating factor.  
 SOURCE Homo sapiens pancreatic cancer cell line: HPC-Y5 cDNA to mRNA, clone: pK027.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 2093)  
 Kojima, T., Oh-beda, M., Hattori, K., Taniuchi, Y., Tamura, M., Ochi, N. and Yamaguchi, N.  
 Molecular cloning and expression of megakaryocyte potentiating factor cDNA  
 J. Biol. Chem. 270 (37), 21984-21990 (1995)  
 MEDLINE 9539496  
 REFERENCE 2 (bases 1 to 10)  
 Kojima, T.  
 AUTHORS Kojima, T.  
 JOURNAL Unpublished (1996)  
 REFERENCE 3 (bases 1 to 2093)  
 Kojima, T.  
 AUTHORS Kojima, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-FEB-1995) to the DDBJ/EMBL/GenBank databases. Tetsuo Kojima, Chugai Pharmaceutical Co. Ltd., Fuji Golemba Research Laboratories; 1-135 Komakado, Golemba-shi, Shizuoka prefecture 412,









QY 520 CGCATCAGGAAGCCAAATGTGACCTGCTCCCGAGGGGGGCTCCCGAGCAGACGGCTG 579  
 Db 543 CGCATCAGGAAGCCAAATGTGACCTGCTCCCGAGGGGGGCTCCCGAGCAGACGGCTG 602  
 QY 580 CTGCTGCGGCTCTGCGCTGCTGGGGGTGTGCGGGGTCTCTGCTGAGGAGGCTGATGTG 639  
 Db 603 CTGCTGCGGCTCTGCGCTGCTGGGGGTGTGCGGGGTCTCTGCTGAGGAGGCTGATGTG 662  
 QY 640 CGGCTCTGAGGAGGCTGCTGGGACCTGCTGGGCTCTGCTGGGCTCTGCTGGGCTCTG 699  
 Db 663 CGGCTCTGAGGAGGCTGCTGGGACCTGCTGGGCTCTGCTGGGCTCTGCTGGGCTCTG 722  
 QY 700 GTGCTGTACCCCGGCTGTGAGCTGCCCGGAGACCCCTGAGCAGAGGACAGAGGCA 759  
 Db 723 GTGCTGTACCCCGGCTGTGAGCTGCCCGGAGACCCCTGAGCAGAGGACAGAGGCA 782  
 QY 760 GCGAGGGGGCTCTGCAAGGGGGGGGAGACCCCTACAGGCCCCCGTGCACATGCTCTG 819  
 Db 783 GCGAGGGGGCTCTGCAAGGGGGGGGAGACCCCTACAGGCCCCCGTGCACATGCTCTG 842  
 QY 820 TCCAGATGAGACGCTCTGCGGGGCTGCTGCGGCTGCTGCGGCTGCGGCTGCGGCTG 879  
 Db 843 TCCAGATGAGACGCTCTGCGGGGCTGCTGCGGCTGCTGCGGCTGCGGCTGCGGCTG 902  
 QY 880 ATCCCGCAGGGCATGTGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 939  
 Db 903 ATCCCGCAGGGCATGTGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 962  
 QY 940 CAGCTGAGGAGGACATCTCTCGGCGCGGCTGCGGCGGAGAGTGGACAAACAGCCCTG 999  
 Db 963 CAGCTGAGGAGGACATCTCTCGGCGCGGCTGCGGCGGAGAGTGGACAAACAGCCCTG 1022  
 QY 1000 CCTTCAGGCAAGAAAGCCCGGAGATAGACAGAGCCCTCATCTTCTACAAAGTGGAG 1059  
 Db 1023 CCTTCAGGCAAGAAAGCCCGGAGATAGACAGAGCCCTCATCTTCTACAAAGTGGAG 1082  
 QY 1060 CTGGAAGCCTGCTGATGATGCGGCGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTG 1119  
 Db 1083 CTGGAAGCCTGCTGATGATGCGGCGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTG 1142  
 QY 1120 CCTTCAGCCTACGAGCAGCTGAGCTGCTTAAGCATTAAGTGAAGTGAAGTGAAGTGA 1179  
 Db 1143 CCTTCAGCCTACGAGCAGCTGAGCTGCTTAAGCATTAAGTGAAGTGAAGTGAAGTGA 1202  
 QY 1180 GGTATACCCGAGTGTGATCAGACACCTGGGCTACCTTCTCTCAAGTGAAGTGAAGTGA 1239  
 Db 1203 GGTATACCCGAGTGTGATCAGACACCTGGGCTACCTTCTCTCAAGTGAAGTGAAGTGA 1262  
 QY 1240 GACATTTGGCAAGTGAATGTACGCTCCCTGAGAGACCTGGAAGGCTTGTGGAAGTGAAG 1299  
 Db 1263 GACATTTGGCAAGTGAATGTACGCTCCCTGAGAGACCTGGAAGGCTTGTGGAAGTGAAG 1332  
 QY 1300 AAAGGCAAGAAATGAGTCTCAGGCTCTGCGGCGGCTCCCAAGAGTGGCCACCTG 1359  
 Db 1323 AAAGGCAAGAAATGAGTCTCAGGCTCTGCGGCGGCTCCCAAGAGTGGCCACCTG 1388  
 QY 1360 ATGACCGGCTTTGTGAAGGGAAGGGGCGAGCTAGACAAAGACACCTAGACACCTGAGC 1419  
 Db 1359 ATGACCGGCTTTGTGAAGGGAAGGGGCGAGCTAGACAAAGACACCTAGACACCTGAGC 1448  
 QY 1420 GCTTTTACCTGCGGTACCTGCTCCCTCAGCCCGGAGGAGTGAAGTGAAGTGAAGTGA 1479  
 Db 1419 GCTTTTACCTGCGGTACCTGCTCCCTCAGCCCGGAGGAGTGAAGTGAAGTGAAGTGA 1478  
 QY 1480 AGCAGCATCTGGGGGCTGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1539  
 Db 1479 AGCAGCATCTGGGGGCTGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1538  
 QY 1540 GTCCCTATCCCAAGGCGGCTGCTTCTCCAGAAATGAAGGAGGAGGAGGAGGAGGAGGAG 1599  
 Db 1539 GTCCCTATCCCAAGGCGGCTGCTTCTCCAGAAATGAAGGAGGAGGAGGAGGAGGAGGAG 1598  
 QY 1600 AAGATCAGTCTCTTCTGCGGTGGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1659

Db 1599 AAGATCAGTCTCTCTCTGCGGTGGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658  
 QY 1660 AATGTGACATGAGCACTTGGGCGGAGCTGATGAAAGTGGGCGGAGGAGTGGGCTGCTGCTG 1719  
 Db 1659 AATGTGACATGAGCACTTGGGCGGAGCTGATGAAAGTGGGCGGAGGAGTGGGCTGCTGCTG 1718  
 QY 1720 ACTGTGCTGAGGTGAGCAAACTTGTGGGAGCCACAGCTGAGGAGGAGGAGGAGGAGGAGGAG 1779  
 Db 1719 ACTGTGCTGAGGTGAGCAAACTTGTGGGAGCCACAGCTGAGGAGGAGGAGGAGGAGGAGGAG 1778  
 QY 1780 CGGACCGGCGGCTGCGGAGTGTGATCTACGCAAGGCGGAGAGCAGACCTGAGACGCTG 1839  
 Db 1779 CGGACCGGCGGCTGCGGAGTGTGATCTACGCAAGGCGGAGAGCAGACCTGAGACGCTG 1838  
 QY 1840 GGGCTGGGGCTACAGAGGCGGAGATCCCAAGCGGCTACCTGCTCTAGACCTAGAGTGCAG 1899  
 Db 1839 GGGCTGGGGCTACAGAGGCGGAGATCCCAAGCGGCTACCTGCTCTAGACCTAGAGTGCAG 1898  
 QY 1900 GAGACCTCTGCGGAGGAGGCGGCTGCTCTAGACCTGAGACCTGCTCTGAGCA 1959  
 Db 1899 GAGGCGCTCTGCGGAGGAGGCGGCTGCTCTAGACCTGAGACCTGCTCTGAGCA 1958  
 QY 1960 CTGCTCTAGCCTCCACCTGCTGCTGAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 2019  
 Db 1959 CTGCTCTAGCCTCCACCTGCTGCTGAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 2018  
 QY 2020 GGGGATCCCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2079  
 Db 2019 GGGGATCCCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2078  
 QY 2080 CGCGCTCAGTAAAGGGAACATGCCCTCGCAGACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2138  
 Db 2079 CGCGCTCAGTAAAGGGAACATGCCCTCGCAGACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2137

RESULT 9  
 AF180951  
 LOCUS  
 DEFINITION  
 Homo sapiens mesothelin/megakaryocyte potentiating factor mRNA,  
 partial cds.  
 AF180951  
 VERSION  
 AF180951.1 GI:6014586.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1195)  
 Scholler,N., Fu,N., Yang,Y., Ye,Z., Goodman,G.E., Hellstrom,K.E.  
 and Hellstrom,I.  
 Soluble member(s) of the mesothelin/megakaryocyte potentiating  
 factor family are detectable in sera from patients with ovarian  
 carcinoma  
 Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11531-11536 (1999)  
 MEDLINE  
 PUBMED  
 99432267  
 10500211  
 2 (bases 1 to 1195)  
 Scholler,N., Yang,Y. and Hellstrom,I.  
 Direct Submission  
 Submitted (26-AUG-1999) Tumor Immunology, PNRI, 720 Broadway Rm.  
 511, Seattle, WA 98107, USA  
 JOURNAL  
 TITILE  
 FEATURES  
 source  
 1. 1195  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 <1. >1195  
 /codon\_start=1  
 /product="mesothelin/megakaryocyte potentiating factor"  
 /protein\_id="AA01409.1"  
 /db\_xref="GI:6014587"  
 /translation="VERTACPSGSKKAREIDESLIFYKKWELEACVDAALLATOMDRVN  
 AIPFYEOLDVLKHKIDELYPQGPESVIOHLGYLFLKMSPEDIRKMNVTSLTAL

LEVNKHEMSPOVALTIDRFVKRGQGLDKDITLDTLTAFLPYGLCSPEBLSSVPS  
 IMAVRDODLCPROLDVLYPKARILAFQMNNGSEVFKIOSEFAGAPTEDLALSOO  
 NMSADLAEFKLRBDALPLTVAEVOXKLEPHVEKAEERHPRVMIILRODDID  
 TLGLGQGTIPNGYIVLDLSVOGRGOAGRAGVEGALSHPICGPRICDNLPR  
 PRTWCSHRRGTAPSLHPGLRPLPCMPPOKSPPOEOARVIVPPOENSHSVNEN  
 MPPADT"

Query Match 45.98; Score 980.8; DB 9; Length 1195;  
 Best Local Similarity 91.18; Pred. No. 3.8e-164;  
 Matches 1108; Conservative 0; Mismatches 2; Indels 106; Gaps 2;

BASE COUNT 235 a 382 c 373 g 205 t  
 ORIGIN  
 Oy 982 GTGAGAGAGACAGCGCTGCTTCAGGGCAAGAGCCCGCGAGATAGACAGAGCTCATC 1041  
 Db 1 GTGAGAGAGACAGCGCTGCTTCAGGGCAAGAGCCCGCGAGATAGACAGAGCTCATC 60  
 Oy 1042 TTCTACAGAAGTGGAGCTGGAAGCCCTGCGTGGATGGCCCTGCTGGCCACAGATG 1101  
 Db 61 TTCTACAGAAGTGGAGCTGGAAGCCCTGCGTGGATGGCCCTGCTGGCCACAGATG 120  
 Oy 1102 GACCGCGTGAAGCCGATCCCTTACCTACGACAGCTGAGCGTCTTAAGCATAACTG 1161  
 Db 121 GACCGCGTGAAGCCGATCCCTTACCTACGACAGCTGAGCGTCTTAAGCATAACTG 180  
 Oy 1162 GATGAGCTCTACCCACAAGGTTACCCGAGTCTGTGATCCAGACGCTGGCTACCTCTC 1221  
 Db 181 GATGAGCTCTACCCACAAGGTTACCCGAGTCTGTGATCCAGACGCTGGCTACCTCTC 240  
 Oy 1222 CTCAAGATGAGCCCTGAGGACATTCGCAAGTGAATGAGCTGCTCGTAGACCTGAG 1281  
 Db 241 CTCAAGATGAGCCCTGAGGACATTCGCAAGTGAATGAGCTGCTCGTAGACCTGAG 300  
 Oy 1282 GCTTGTCTGAAGTCGACAAAGGGCAGAAATGAGTCTCTAGGCTCCTCGCGGCCCTC 1341  
 Db 301 GCTTGTCTGAAGTCGACAAAGGGCAGAAATGAGTCTCTCTCTCTCTCTCTCTCTCT 339  
 Oy 1342 CCACAGGTGGCCACCTGATGACCGCTTTGTGAAGGAGAGGGCCACCTAGACAAAGAC 1401  
 Db 340 ---CAGTGGCGCCACCTGATGACCGCTTTGTGAAGGAGAGGGCCACCTAGACAAAGAC 396  
 Oy 1402 ACCCTAGACACCTGACCGCTTCTACCTGGTAACTGTGCTCCCTCAGCCCGAGAGAG 1461  
 Db 397 ACCCTAGACACCTGACCGCTTCTACCTGGTAACTGTGCTCCCTCAGCCCGAGAGAG 456  
 Oy 1462 CTGAGCTCCGCGCCCGCCAGAGCATGTGGCGGTCAAGGCCCGACAGCTTGACACGT 1521  
 Db 457 CTGAGCTCCGCGCCCGCCAGAGCATGTGGCGGTCAAGGCCCGACAGCTTGACACGT 516  
 Oy 1522 GACCAAGGAGCGTGAAGCTCTATCCCAAGGCCCGCTTGTCTCAGACATGAAC 1581  
 Db 517 GACCAAGGAGCGTGAAGCTCTATCCCAAGGCCCGCTTGTCTCAGACATGAAC 576  
 Oy 1582 GGGTCCGAATATTCTGTGAAGATCCAGTCTCTGGTGGGGCCCGACAGAGATTG 1641  
 Db 577 GGGTCCGAATATTCTGTGAAGATCCAGTCTCTGGTGGGGCCCGACAGAGATTG 636  
 Oy 1642 AAGGCGCTCATATGACAGAAATGTAGATGACTTGGCCACGTCATGAACCTGCGAGC 1701  
 Db 637 AAGGCGCTCATATGACAGAAATGTAGATGACTTGGCCACGTCATGAACCTGCGAGC 696  
 Oy 1702 GATCGGTGTGCGCTGTGACTGTGCTGAGTGCAGAAATCTTGGGAGCCCGACAGTGAG 1761  
 Db 697 GATCGGTGTGCGCTGTGACTGTGCTGAGTGCAGAAATCTTGGGAGCCCGACAGTGAG 756  
 Oy 1762 GGCTTGAAGCGGAGAGCGGACCGCGGTGCGGAGCTGATCTTAAGCGAGCGGAG 1821  
 Db 757 GGCTTGAAGCGGAGAGCGGACCGCGGTGCGGAGCTGATCTTAAGCGAGCGGAG 816  
 Oy 1822 GAGCAGCTGAGACAGCTGGGGCTGAGGGCTACAGGGCGGACATCCCAACGGCTACCTGTC 1881  
 Db 817 GAGCAGCTGAGACAGCTGGGGCTGAGGGCTACAGGGCGGACATCCCAACGGCTACCTGTC 876

Oy 1882 CTAGACCTCAGCGTGA----- 1898  
 Db 877 CTAGACCTCAGCGTGAAGTGGGCGGGCGGCCAGGCCAGGCGTGGGGCAGAGCTGG 936  
 Oy 1899 -----AGAGACCTCTCGGGGAGCGC 1919  
 Db 937 GCGGTGAGGTGGGCGCTGTGAGTCACCCCTCTCTCTGTGAGAGGCGCTCTCGGGAGCGC 996  
 Oy 1920 CTGCTCTAGAGACCTGGACCTGTTCTACCGTCTGTGACATGCTCTCAGCTCGACCT 1979  
 Db 997 CTGCTCTAGAGACCTGGACCTGTTCTACCGTCTGTGACATGCTCTCAGCTCGACCT 1056  
 Oy 1980 GCGCTGAGGGCGCCACATCCCTTGGCGGCGCCAGCGCTGCGGGATCCCGCGTGGAGC 2039  
 Db 1057 GCGCTGAGGGCGCCACATCCCTTGGCGGCGCCAGCGCTGCGGGATCCCGCGTGGAGC 1116  
 Oy 2040 GAGCAGCGAGGATGATCCCGTTCACCCCAAGAGAACTGCGCTCAGTAAAGGGAGC 2099  
 Db 1117 GAGCAGCGAGGATGATCCCGTTCACCCCAAGAGAACTGCGCTCAGTAAAGGGAGC 1176  
 Oy 2100 ATGCCCCCTGACACA 2115  
 Db 1177 ATGCCCCCTGACACA 1192

## RESULT 10

D86370 2135 bp mRNA ROD 07-FEB-1999  
 LOCUS  
 DEFINITION mouse mRNA for megakaryocyte potentiating factor, complete cds.  
 ACCESSION D86370.1 GI:1408516  
 VERSION  
 KEYWORDS megakaryocyte potentiating factor.  
 SOURCE Mus musculus tissue\_l1b:Lung cDNA to mRNA.  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Kojima,T., Taniguchi,Y., Hattori,K. and Oh-eda,M.  
 TITLE Submitted (03-JUL-1996) to the DDBJ/EMBL/Genbank databases. Tetsuo  
 JOURNAL Kojima, Chugai Research Institute for Molecular Medicine, Inc.,  
 Cytokine Research Program; 153-2, Nagai, Mihari, Ibaraki 300-41,  
 Japan (E-mail:kojimatemb.infoweb.or.jp, Tel:0298-30-6211,  
 Fax:0298-30-6270)  
 2 (bases 1 to 2135)

REFERENCE  
 AUTHORS Kojima,T., Taniguchi,Y., Hattori,K. and Oh-eda,M.  
 TITLE mouse Megakaryocyte Potentiating Factor cDNA  
 JOURNAL unpublished (1996)  
 REFERENCE  
 AUTHORS Kojima,T., Oh-eda,M., Hattori,K., Taniguchi,Y., Tamura,M., Ochi,N.  
 and Yamaguchi,N.  
 TITLE Molecular cloning and expression of megakaryocyte potentiating  
 factor cDNA  
 JOURNAL J. Biol. Chem. 270 (37), 21984-21990 (1995)  
 MEDLINE  
 FEATURES  
 source Location/Qualifiers

1..2135  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /tissue\_lib="Lung"  
 109..255  
 109..1986  
 /function="stimulates the megakaryocyte colony forming  
 activity of murine Interleukin-3 in mouse bone marrow cell  
 culture"  
 /codon\_start=1  
 /product="megakaryocyte potentiating factor"  
 /protein\_id="BAI3077.1"  
 /db\_xref="GI:1408517"  
 /translation="MALPAPRPLLSGSGSPICRSRLLLLSGLMIPRLDTOTTKTSO  
 EATLHAVNGADPASPGLGVLCEEVSDISMEGAKLAIAVVRKNTLKGHOL  
 CLARRLRRLTDELNALPDLILLFLNPAWPFQOQCAHFPSLISKRVNDVLPRLSLE

sig\_peptide  
 CDS



QY 1971 CTCACCCCTGGCTGAGGGCCCCACATCCCT-----TGTGGCCCCAGCCCTGCTGGGGAT 2025  
 Db 1971 TTAAAGGCTGAGCTGAGACACACACCTCGAAGGCTCTGCTCCAGCTACTGAGGG-- 2028  
 QY 2026 CCCCCCTGGCCAGAGCAGCAGGGTATCCCGTTCCACCCCAAGAGACTGCGCCT 2085  
 Db 2029 -CCCTTTCAGACAGAGGCTACAGGGCTCATTTGCCAAAGTTTGAGAGCTCTTGAAC 2087  
 QY 2086 CAGTAACGGGACATGCCCCCTGCAGACAAACAAAAA 2133  
 Db 2088 CAATTAACAGTGGCATATGCTCCTTGAACAAAAA 2135  
 RESULT 11  
 LOCUS DB7351 2099 bp mRNA ROD 29-SEP-2000  
 DEFINITION Rattus norvegicus mRNA for mesothelin, complete cds.  
 ACCESSION DB7351.1 GI:10129877  
 VERSION DB7351.1  
 KEYWORDS Erc; mesothelin.  
 SOURCE Rattus norvegicus (strain:Eker rat) renal carcinoma cell\_line:ERC33  
 ORGANISM Rattus norvegicus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (sites)  
 Yamashta, Y., Yokoyama, M., Kobayashi, E., Takai, S. and Hino, O.  
 Mapping and determination of the cDNA sequence of the Erc gene  
 preferentially expressed in renal cell carcinoma in the Tsc2 gene  
 mutant (Eker) rat model  
 Blochem. Biophys. Res. Commun. 275 (1), 134-140 (2000)  
 2 (bases 1 to 2099)  
 Hino, O.  
 Direct Submission  
 Submitted (23-AUG-1996) to the DDBJ/EMBL/GenBank databases, Oki  
 Hino, Cancer Institute, Department of Experimental Pathology;  
 Kami-Ikebukuro 1-37-1, Toshima-Ku, Tokyo 170-8455, Japan  
 (E-mail:ohino@ims.u-tokyo.ac.jp, Tel:03-5394-3815,  
 Fax:03-5394-3815)  
 FEATURES  
 source Location/Qualifiers  
 1. 2099  
 /organism="Rattus norvegicus"  
 /strain="Eker rat"  
 /db\_xref="taxon:10116"  
 /cell\_line="ERC33"  
 /cell\_type="renal carcinoma"  
 1. 90  
 91..1968  
 /gene="Erc"  
 91..1968  
 /gene="Erc"  
 /note="homologue of megakaryocyte potentiating factor  
 (MPF)"  
 /codon\_start=1  
 /product="mesothelin"  
 /protein\_id="BA013512.1"  
 /db\_xref="GI:10129877"  
 /translation="MALPTAQLPLGSGSPICSRPLLLLSLGLPDLTOTRTISO  
 EALHVAVTGTPASLPTGLFGLTDEVSGSLWGHAKELAMAVRKNTVLOVHQUR  
 CLARRLPKHLTNEELALPLDLLFLNPAHPGOACAHFSLISKANVLPKRSLE  
 RORLLTALCKOCQYGVQSETPDARALGACLDLPGFVAKSSSEVLPMVLRGSGPLD  
 OGAKAREVLRSGARAYGPSTWSSTLALQGLVVLDESIVHSIPKDYITMGLG  
 ISREPSRLGSKMTYTHPRFRDTEOKACPRGKPEKNVDENLIFYONNELFACVDCITL  
 AGOMDLYNEPTFYEQSLIFKRLDKTYPGQPSLSLKQLGHFFRYSPEDIRMYNT  
 SPDVNTLLVSKQKMDAOLVIALVACYLVGSGKLDLDIVKALDNIPLSLDCSPD  
 LHAIPSSVMVLVGLHLDLDCOSRHLGILYKACSAFQNVSGLEVEFIRFLGASPE  
 DLRALSGHNSMDIATFKKLOVALVGLSVAEQKLGLPHIGDKTEEDKSPVDMFL  
 ROOKDSDSLGLGIGIPNGYILIDENVEAFSSGAPLLGPVFAMIDALLSLRL  
 S"  
 1969..2099  
 2074..2079  
 3'UTR  
 polyA\_signal

BASE COUNT 465 a 588 c 585 g 461 t  
 ORIGIN  
 Query Match 42.2%; Score 902.8; DB 10; Length 2099;  
 Best Local Similarity 67.9%; Pred. No. 2e-150;  
 Matches 1400; Conservative 0; Mismatches 612; Indels 50; Gaps 8;  
 QY 69 GGTGCTCCCTCCCTGGATCTACAGACATGAGCCCTG-CAGCGGTGAGCCCTGTT 127  
 Db 60 GGAACCTGAGAACCATTTGTTATCCACAGACATGAGCCCTTGCACAGCCCAACCCCTGCT 119  
 QY 128 GGTCTGTGGGGA-----CCGCCCCGAGAGCTCTCTGTTCCGTCTTCAAGCTCGG 179  
 Db 120 GGGGTCTGTGGAGGCCCATCTGAGCCGAGCTTCTACTGCTTCTCCCTAAGCTTGG 179  
 QY 180 ATGGGTGATCCCGGAGAGCCCTGAGCTGAGAGACAGGAGCGAGTGTGCCCCCTGG 239  
 Db 180 GTGGTTGCCACTTCTGCGAGAGACTACAGACAAAGCAGGAGAGCGGCACTTCTCA 239  
 QY 240 GGGAGTCTGACACACCCCATTAACATTTCCAGGCTCTCCCTGGCCAACTGTTGGCTT 299  
 Db 240 TGCTG---TGACCGGACACCTTGACTTGCCAGTCTTCCACAGAGCTCTTTTGGCTT 296  
 QY 300 CCGGTGTCGGAGGTGTCGGGCTGAGCAGCAGAGCTGTCCGGAGCTGCTGTGGCCTT 359  
 Db 297 CACGTGTATAGAGTATCTGGCTTACAGCTGAGCAGCCCAAGAGAGCTGCTATGGCTGT 356  
 QY 360 GGCACAGAGATGTCAAGCTCTCACAGAGACAGCTGCGTGTCTGCTCACCGGCTTC 419  
 Db 357 GAGACAGAGAAATATCGTCTCAAGTACATCAGCTGCGTGTCTGCGCCCTGCTCC 416  
 QY 420 TGAGC-----CCCCCGGAGACCTGGAGCAGCCCTCCATTGGACCTGCTATTCTTAA 473  
 Db 417 TAAGCACTCACACAGAGAGACTGATGATCTCTCCACTGAGACCTGCTCTTCTTAA 476  
 QY 474 CCGAGATCGTTCTCGGGGCCCCAGAGCCCTGCACCCGTTTCTTCCCATCAGCAAGGC 533  
 Db 477 TCCAGCATTTGTTCCGGGGCAAGAGGCTTGTGCCACTTCTTCCCTCATCTTAAAGC 536  
 QY 534 CAATGTGACCTGCTCCCGAGAGGGGGCTCCGAGCGACAGAGCGCTGCTGCGGCTCT 593  
 Db 537 CAATGTAATTAATCTCTCCACGAGATCTCTGAGCGCCAGAGGCTGCTGACCGGGCTCT 596  
 QY 594 GGCCCTGAGGCTGAGGGGGCTCTCTCTGAGCAGGCTATGAGCGGGCTGAGGAGG 653  
 Db 597 GAATGCGAGGCTGTGATGATTTCAAGTGAAGACAGGATGACGGGCTCTCGAGG 656  
 QY 654 CCTGGCTTGCAGCTGCTGAGGCGCTTGTGTGCGGAGTGGCCGAAGTGTCTAACCCG 713  
 Db 657 CCTGGCTGTGACCTGCTGAGGAAATTCGTGGCCAAATCTTGGAAAGTCTCTCCCTG 716  
 QY 714 GCTGTGAGCTGCCCCGAGACCCCTGAGACAGGACAGCAGAGGAGCGAGCGGCTCT 773  
 Db 717 GCTGCAAGATGCGGAGAGCCTGTGAGCAGGACAGGCAAGGCTGTCAAGGAGTCT 776  
 QY 774 GAGGGGCGGGGAGACCCCTACGGGCCCCGCTGCACATGCTGTCTCAGATGAGAGC 833  
 Db 777 GAGAGGTGAGAGAGGCCCTTATGCTGCCCATGCACAGTGTAGCTCTCACACCTGAGTC 836  
 QY 834 TGTGGGGGCTGCTGCGCGTGTGCGGCGAGCCATCATCCGAGCAATCCCGAGGCGAT 893  
 Db 837 CCTGAGAGGTTTGGCTGAGTGTGGATGATGATTCATTTCCACAGCATCTCTTAAGAGAT 896  
 QY 894 CGTGGCGCGGAGGAGAGCGCTCTCTCTCGAGACCATCTCGGGGGGAGCGCTGAAGCGAG 953  
 Db 897 TATCACTAATAGCTGCGCAAGCATCTCCALAGAGCCCTCCAGGCTGGGCTTAAGTGAC 956  
 QY 954 CATCTCCGGCGGGCTGTCGGCGGGAAGTGCAGACAGACACCTGCTCTTGAAGCAAGAA 1013  
 Db 957 TGTACACACCCCAAGGTTCCCGGGGACACACAGAACGAAACCTGCTCCAGAGGAAGA 1016  
 QY 1014 GGCCCGGAGATAGACGAGACCTCATCTCTTCTACAAAGATGGAGCTGGAAGCCTCGCT 1073

Db	1017	GCCTAACGTGGTGGATGAATAAACCATTCTTTCTACCAAGATTGGAGCTGGAGGCTTGCT	1076
Oy	1074	GGATGCGGCCCTGCTGGCCACCCAGATGGACGCGGTGAAGCCATCCCTTACCTACGA	1133
Db	1077	CGATGGTACCCCTGCTGGCCCGCCAGATGGACCTGTGAATGAAATTCCTTTACTACGA	1136
Oy	1134	GCACCTGAGAGTCCCTAAAGCATTAACCTGGATGAGCTCTACCCAGAAAGTTACCCGAGTC	1193
Db	1137	GCACCTGACGATCTTTCAGACCAAACTGGACAAAGACTTACCCACAAGGCTATCCGAGTC	1196
Oy	1194	TGTGATCCAGACACTGGGCTACCTCTTCTCAAGATGAGCCCTCGAGACATTCGCAATG	1253
Db	1197	CCGTATCAAGAGCTGGGCCCATTTCTTCAGATACGTTAGCTCCGAGAGCATCCGGCAATG	1256
Oy	1254	GAAATGTACGCTCCCTGGAGACCCCTGAAGGCTTTGCTTGAAGTGCACAAAGGCGAGAAAT	1313
Db	1257	GAAATGTACCTTACACACACACAGTGAATCCTGCTTAAAGTGCACAAAGGCGAGAAAT	1316
Oy	1314	GAGTCTCAGGCTCTCTGGGCGGCCCTCCACAGGTGGCGCACCTCGATGACGCTTGTT	1373
Db	1317	GGATGCT-----CAGTGATGTGCTTGCTGCTGCTATCT	1352
Oy	1374	GAAAGGAAGGGGCCAGCTAGACAAAGACACCTAGACACCTGACCGCTTCTACCTGG	1433
Db	1353	TGCGGAGAGAGGCGCAAGCTGGACGAGACATAGTAAAGCCCTCGGACAAACATCCCTTAAG	1412
Oy	1434	GTACCTGTGCTCCCTTCAGCCCGGAGAGGCTGAGCTCCGCGCCCGCAGAGAGATCTGGGC	1493
Db	1413	TTACTATATGATCTTACGCCCCCAGAGATCTGCACGCTATACCTCTCAGTGTATGTGCT	1472
Oy	1494	GGTCAGAGCCCCAGAGACTGTGACACGTTGACCCCAAGCAGCTGGACGTTCTCTATCCAA	1553
Db	1473	GGTTGGGCTCCATGACTCTGGACAAGTGCAGCGACAGGCGCATCGGTATCTCTATACAA	1532
Oy	1554	GGCCCGGCTTGTCTCCAGAAACATGAAGCGGTCCGATCTCTGTGAAGATCCAGTCTT	1613
Db	1533	GGCTGCTCAGCCTTCCAGAAAGGTGAGCGGGCTGTGATCTTTGAGAAATTCAGGACATTT	1592
Oy	1614	CTGTGGTGGGGCCCCCAGGAGGATTTGAAGGGCGCTCAGTCAAGCAAAATGTGAGCATGA	1673
Db	1593	TCTGTGGGGGCTTCCAGGAGAGACCTCGGGGGCTTCAGCCAGCAAAATGTGAGTATGA	1652
Oy	1674	CTTGGCCACGTTTATGAAGCTGGGACGGAATGGGTGCTGCTTGACTGTGCTGAGAGT	1733
Db	1653	CATGAGCCATTTCAAGAAAGCTGAGGTGATGCTGTGGGGCTGAGTGTGCTGAGAGT	1712
Oy	1734	GCAAAACCTTCTGGAGACCCACGCTGGAGGGCCTGGAAGGCGGAGAGCGGACCGCCGCT	1793
Db	1713	ACAGAAACTTCTAGGGCCACACATTTGGGACCTGGAAGATGAGAGAGATAAAGCCGT	1772
Oy	1794	CGGGAGCTGATCTCTACGCGAGCGGACAGGACGACCTGACACGCTGGGGCTGGGGCTACA	1853
Db	1773	CCGGAGACTGGCTCTTCCGACAGCAGCAAGAAAGCCTGGACACTCGGTTTGGGACTTCA	1832
Oy	1854	GGGGGGATCCCAAGGGGCTACCTGTGCTCAAGCCTCAGCGTGAAGGAGAGACCTCTCGGG	1913
Db	1833	GGGGGGATCCCAAGGGGCTACCTGTGCTCAAGCCTCAGCGTGAAGGAGAGACCTCTCGAG	1892
Oy	1914	GAGCGCCCTGCTCTTAGAGACTGGAGCTGTTCACAG---TCTGGGAGCTGCTTAGC	1970
Db	1893	TGAGAGCCCACTCCTTGGGCGCAGAGATTTGTGTGTGATGGAATTCACGCTCTGCTACG	1952
Oy	1971	CTTCACCCCTGCTGAGGCGCCCATCTC---CTTGCTGGCCCCAGCCCTGCTGGGATC	2026
Db	1953	TTTAAAGTACTGAGCTGAGACCAACCATCTTAAGGCTCTGCTGCCAGCTCTATTGTGAGGC	2012
Oy	2027	CCG---CGCTGGGCGAGGACAGGACGAGGGTATCTCCGTTGCACCCCAAGGAACCTGGGCT	2085
Db	2013	CCCATCTTTGACAGGAGGAGATACAGAGGCTCATTTGCCAAATTTTGAAGATTTCTTGAACC	2072
Oy	2086	CAGTAAACGGGAAACATGCCCCC	2107

DB	2073	CAATTAACGTCGATGTGCC	2094
RESULT	12		
LOCUS	HS335H7		
DEFINITION	Human DNA sequence from clone LA16-335H7 on chromosome 16 Contains the MSLN (mesothelin) gene, GSS and CpG islands, complete sequence.		
ACCESSION	AL031258		
VERSION	AL031258.12		
KEYWORDS	HTG; CpG island; mesothelin; MSLN.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 37307)		
TITLE	Cobley, V.		
JOURNAL	Direct Submission		
COMMENT	Submitted (17-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humdunye@sanger.ac.uk On Apr 22, 2001 this sequence version replaced g1:9798442. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats: all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL, SW.; SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep/LA16-335H7 is part of a clone contig from the tip of the short arm of chromosome 16 spanning 2Mb of pl3.3 (Higgs D.R., Flint J., Daniels R., MRC Molecular Haematology Unit, Institute of Molecular Medicine, Oxford (unpublished)), and is from the Los Alamos, flow sorted human Chromosome 16 libraries constructed by Norman Doggett (unpublished). VECTOR: scos-1 IMPORTANT: This sequence is not the entire insert of clone LA16-335H7. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone LA16-335H7 is at 1 in this sequence. The true left end of clone LA16-331D is at 37208 in this sequence. The true right end of clone LA16-444C9 is at 6146 in this sequence. Location/Qualifiers 1. .37307 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="16" /clone="LA16-335H7" /clone_1fb="LA16" 2. .199 /note="THEIC repeat: matches 99. .302 of consensus" repeat_region 200. .510 /note="Alusq repeat: matches 1. .312 of consensus" 511. .581 /note="THEIC repeat: matches 302. .371 of consensus" 741. .1033 /note="Alub repeat: matches 1. .301 of consensus" 1620. .1931 /note="Alusx repeat: matches 1. .312 of consensus" 1991. .2128 /note="FLAM_C repeat: matches 1. .133 of consensus" 2332. .2626 repeat_region		

```
/note-"AluSg repeat: matches 1. .295 of consensus"
2638. .2929
/note-"AluSg repeat: matches 5. .298 of consensus"
repeat_region
2961. .3124
/note-"AluSg repeat: matches 137. .300 of consensus"
repeat_region
3125. .3419
/note-"AluY repeat: matches 1. .295 of consensus"
repeat_region
4098. .4179
/note-"MIR repeat: matches 82. .166 of consensus"
repeat_region
4269. .4320
/note-"L2 repeat: matches 2449. .2501 of consensus"
repeat_region
5415. .5668
/note-"AluSx repeat: matches 1. .303 of consensus"
repeat_region
5677. .5905
/note-"LIMB4 repeat: matches 5905. .6130 of consensus"
repeat_region
5906. .6200
/note-"AluSx repeat: matches 5. .301 of consensus"
repeat_region
6201. .6313
/note-"LIMB4 repeat: matches 5792. .5905 of consensus"
repeat_region
6317. .6589
/note-"MER49 repeat: matches 5. .285 of consensus"
repeat_region
6674. .6793
/note-"MER4D repeat: matches 411. .536 of consensus"
repeat_region
6794. .7096
/note-"AluSg repeat: matches 1. .304 of consensus"
repeat_region
7097. .7292
/note-"MER4D repeat: matches 536. .981 of consensus"
repeat_region
7303. .7590
/note-"AluSx repeat: matches 1. .290 of consensus"
repeat_region
7596. .7888
/note-"LIMB4 repeat: matches 5518. .5795 of consensus"
repeat_region
7895. .7994
/note-"LIM4 repeat: matches 5490. .5589 of consensus"
repeat_region
8004. .8297
/note-"AluSx repeat: matches 1. .294 of consensus"
repeat_region
8288. .8433
/note-"AluSg/x repeat: matches 1. .136 of consensus"
repeat_region
8445. .8510
/note-"LIM4 repeat: matches 6252. .6322 of consensus"
repeat_region
8511. .8796
/note-"AluY repeat: matches 1. .298 of consensus"
repeat_region
8797. .8894
/note-"LIM4 repeat: matches 6172. .6252 of consensus"
repeat_region
8967. .9569
/note-"LIMB2 repeat: matches 5402. .5988 of consensus"
repeat_region
9571. .9678
/note-"LIM4 repeat: matches 5379. .5462 of consensus"
repeat_region
9679. .9969
/note-"AluY repeat: matches 2. .288 of consensus"
repeat_region
9970. .10029
/note-"LIM4 repeat: matches 5314. .5379 of consensus"
repeat_region
10031. .10076
/note-"LIMB8 repeat: matches 6126. .6173 of consensus"
repeat_region
10097. .10229
/note-"AluSg/x repeat: matches 1. .135 of consensus"
repeat_region
10230. .10530
/note-"AluSx repeat: matches 1. .302 of consensus"
repeat_region
10747. .10844
/note-"LIMB2 repeat: matches 5162. .5254 of consensus"
repeat_region
10875. .11025
/note-"LIMB2 repeat: matches 4946. .5097 of consensus"
repeat_region
11048. .11336
/note-"AluY repeat: matches 1. .295 of consensus"
repeat_region
11372. .11499
/note-"AluY repeat: matches 10. .132 of consensus"
repeat_region
11500. .11803
/note-"AluY repeat: matches 2. .305 of consensus"
repeat_region
11805. .12112
/note-"AluSx repeat: matches 1. .308 of consensus"
repeat_region
12113. .12218
/note-"AluY repeat: matches 3. .10 of consensus"
repeat_region
12820. .13201
/note-"match: GSS: Em:AQ166427"
```

---

```
repeat_region
13363. .13521
/note-"3 copies 53 mer 88% conserved"
repeat_region
13670. .13756
/note-"MIR repeat: matches 103. .191 of consensus"
repeat_region
13902. .14202
/note-"MER54B repeat: matches 310. .618 of consensus"
mRNA
join(15529. .15700,16168. .16211,16577. .16626,16958. .17078,
17582. .17661,17842. .17971,18045. .18238,18462. .18552,
18626. .18725,18994. .19172,19270. .19425,19579. .19745,
19820. .19947,20311. .20405,21314. .21500,21583. .21668)
/gene-"MSLN"
/note-"match: cDNAs: Em:U40434 Em:D49441 Em:E07943
Em:D86370"
/evidence-not experimental
/product-"C335H7.1 (mesothelin)"
15529. .21800
/gene-"MSLN"
join(15616. .15700,16168. .16211,16577. .16626,16958. .17078,
17582. .17661,17842. .17971,18045. .18238,18462. .18552,
18626. .18725,18994. .19172,19270. .19425,19579. .19745,
19820. .19947,20311. .20405,21314. .21500,21583. .21668)
/gene-"MSLN"
/note-"match: proteins: Tr:Q61468 Tr:Q14859"
/codon_start=1
/evidence-not experimental
/product-"C335H7.1 (mesothelin)"
/protein_id="CAC37289.1"
/db_xref="GI:13751645"
/translation="MALPFRPLGSGCTPALSGSLFLFLSLGVQPSRLAETGQE
AAPLDGLVAPNPVSSLSPPQLGFPFAEVSGLSTVERLAVAAKQNKVLTEDIR
CLAHRLSEPPEDDALPLDLFLNPDPAPGPAQCFESRIRKANVDLPRAPERO
RLPLPALACGVSGLSLSEADVRALGDLGRFVAESAAYLLPLVSCPLDD
QOEARALOGGEPPEPPEPSTWSVTMDARGLPLVGOIISIPGIVAWRORS
RDSRWPEPRTIIPRRREVETACSGKARIDSLFVKKMLFAVDAAIAT
QMDRVNIPPTTECLDVLRKRLDELTPRGPESTIONHGLFLPKMSPEDIRKNVSL
ETLKALEVKKGHMSFQAPRRPLPOVALIDRVKRGQLDDTLTLPALPGYLC
SLSEPLEVSPSSIVAWRPDDLTCPDPLDVLVPRARAFNMNSELVFIQSL
GGAFTEDLKALSOQNSMDLATPMKLTDAVLPLTVAVQKLLGPHREGKAEARHP
VRDMILRRDDLDITGLGQGGIPNCTVLDLSMQALSPCLLGLPGVLVLL
LASTIA"
16704
/gene-"MSLN"
/note-"Random repeat. forced join. Restriction digest data
(ECORI, HINCII, SmaCI) suggest 400 bases are missing."
complement(16716. .17130)
/note-"match: GSS: Em:AQ187418"
17850. .18744
/gene-"MSLN"
/note-"Cpg island"
/evidence-not experimental
18858. .18989
/note-"MIR repeat: matches 21. .147 of consensus"
misc_feature
20288. .20986
/gene-"MSLN"
/note-"Cpg island"
/evidence-not experimental
20410. .21099
/note-"15 copies 46 mer 67% conserved"
repeat_region
20436. .21077
```

---

```
Query Match 14.4%; Score 307.6; DB 9; Length 37307;
Best Local Similarity 82.4%; Pred. No. 1.9e-45;
Matches 402; Conservative 0; Mismatches 4; Indels 82; Gaps 1;

QY 1710 GCTGCCGTTGACTGCTGAGTGACACAACTTCTGGACCCACGTGAGGCGCTGAA 1769
|||
DB 21310 GCACCCGTTGACTGCTGAGTGACAAACTTCTGGACCCACGTGAGGCGCTGAA 21369

QY 1770 GCGGAGAGGCGGACCGCCGGTGCGGACTGATCTTACGGACGCGGACGAGACT 1829
|||||
DB 21370 GCGGAGAGGCGGACCGCCGGTGCGGACTGATCTTACGGACGCGGACGAGACT 21429

QY 1830 GGACACGCTGGGCTGAGGCTACAGGGCGGCATCCCAACGCGTACCTGTCTAGACT 1889
```











	repeat_region	23/10/2000"	/rpt_family="Simple_repeat"	complement(35112..35278)	/note="MEN3; Repeatmasker predicted 23/10/2000"
mRNA	gene	joh(35531..36612,36910..37537,39262..39311,39800..40655)	/rpt_family="DNA" joh(35531..36612,36910..37537,39262..39311,39800..40655) /gene="Rjd1"	homology support for exon1"	HS.121190 & 201275; no
CDS	gene	35531..41726 /gene="Rjd1"	homology support for exon1"	HS.121190 & 201275; no	
	Query Match	Best Local Similarity 14.4%; Score 307.6; DB 9; Length 256073;	Mismatches 4; Indels 82; Gaps 1;		
Y 1710	GCTGCGCTTGACTGGCTGGGTGCAGAACTTGTGGACCCCAAGGAGGAGGCTGAA	1769			
Db 240668	GCAGCCGTTGACTGTGGCTGAGTGCATAAATTGTGGACCACCTGGAAGGCTTGA	240667			
QY 1770	GCGGAGGAGGCGCACGCCCGGTGCGGAGCTGATTCCTACGGCAGCGCAGACACT	1829			
Db 240668	GCGGAGGAGGCGCACGCCCGGTGCGGAGCTGATTCCTACGGCAGCGCAGACACT	240727			
QY 1830	GGACACCTGTGGGCGCTGAGGGCTACAAGGGCGCATCCCAACGGCTACCTGTCTTAGACT	1889			
Db 240728	GGACACCTGTGGGCGCTGAGGGCTACAAGGGCGCATCCCAACGGCTACCTGTCTTAGACT	240787			
QY 1890	CAGCGTGCA-----	1898			
Db 240788	CAGCATGCAAGGTGGGGGGGGGGCGCCACAGGCGCTGGGGCACAAGCTGGGGGGCTGGA	240847			
QY 1899	-----AGAACCCCTCTCGGGGAGACGCCCTGCTTC	1927			
Db 240848	GGTGGCGCTCTGACTCACCCCTCTCTGTAGAAGGCCCTCTCGGGGAGACGCCCTGCTTC	240907			
QY 1928	TAGGACCTGGACCTGTTGTCAACCGTCTGGGCACTGCTCTTAGCTTCACACCTCGGCGCTGAG	1987			
Db 240908	TAGGACCTGGACCTGTTGTCAACCGTCTGGGCACTGCTCTTAGCTTCACACCTCGGCGCTGAG	240967			
QY 1988	GGCCCCACTCTCCCTGCTGCTGGCCCCAGCCCTGCTGGGGGATCCCGCCTGGCCAGAGACAGGC	2047			
Db 240968	GGCCCCACTCTCCCTGCTGCTGGCCCCAGCCCTGCTGGGGGATCCCGCCTGGCCAGAGACAGGC	241027			
QY 2048	ACGGGTGATCCCGGCTTCACACCCCAAGAGAACTCGGGCTAGTAACGGGAACATGCCCCC	2107			
Db 241028	ACGGGTGATCCCGGCTTCACACCCCAAGAGAACTCGGGCTAGTAACGGGAACATGCCCCC	241087			
QY 2108	TGCAGACA 2115				
Db 241088	TGCAGACA 241095				
RESULT 15	LOCUS	AC092381/c			
DEFINITION	Homo sapiens chromosome 16 clone RP11-728H8, WORKING DRAFT				
ACCESSION	AC092381				
VERSION	AC092381.1 GI:14589570				
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. DOE Joint Genome Institute. Sequencing of Human Chromosome 16 Unpublished				

REFERENCE	(bases 1 to 155770)
AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT	-----Genome Center Center: Joint Genome Institute Center Code: JGI Web site: <a href="http://www.jgi.doe.gov">http://www.jgi.doe.gov</a> ----- Project Information Center Project Name: 1574232 Center clone name: RPC1-11_728H8 ----- Summary Statistics Consensus quality: 13310 bases at least Q40 Consensus quality: 145364 bases at least Q30 Consensus quality: 148096 bases at least Q20 Estimated insert size: 154000; agarose-fp estimation Estimated insert size: 153570; sum-of-contigs estimation Quality coverage: 5.01 in Q20 bases; agarose-fp estimation Quality coverage: 5.03 in Q20 bases; sum-of-contigs estimation. NOTE: This is a 'working draft' sequence. It currently * consists of 23 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. 1 1258: contig of 1258 bp in length 1259 1358: gap of unknown length 1359 2861: contig of 1503 bp in length 2862 2961: gap of unknown length 2962 5359: contig of 2398 bp in length 5360 5459: gap of unknown length 5460 6559: contig of 1100 bp in length 6560 6659: gap of unknown length 6660 9446: contig of 2787 bp in length 9447 9546: gap of unknown length 9547 12043: contig of 2497 bp in length 12044 12143: gap of unknown length 12144 14853: contig of 2710 bp in length 14854 14953: gap of unknown length 14954 17685: contig of 2732 bp in length 17686 17785: gap of unknown length 17786 19312: contig of 1527 bp in length 19313 19412: gap of unknown length 19413 21856: contig of 2444 bp in length 21857 21956: gap of unknown length 21957 24847: contig of 2891 bp in length 24848 24947: gap of unknown length 24948 27862: contig of 2915 bp in length 27863 27962: gap of unknown length 27963 33524: contig of 5562 bp in length 33524 33624: gap of unknown length 33525 33625: contig of 4877 bp in length 33625 38501: contig of unknown length 38501 38601: gap of unknown length 38602 46486: contig of 7885 bp in length 46487 46586: gap of unknown length 46587 54839: contig of 8253 bp in length 54840 54939: gap of unknown length 54940 65287: contig of 10348 bp in length 65288 65387: gap of unknown length 65388 76253: contig of 10866 bp in length 76254 76353: gap of unknown length 76354 88461: contig of 12108 bp in length 88462 88561: gap of unknown length 88562 102317: contig of 13656 bp in length 102318 102317: gap of unknown length 102318 118382: contig of 16075 bp in length 118383 118492: gap of unknown length 118393 133959: contig of 15467 bp in length 118493

\* 133960 134059: gap of unknown length  
\* 134060 155770: contig of 21711 bp in length.  
FEATURES  
source  
1. 155770  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-728H8"  
/clone.lib="RPII human BAC library 11"

BASE COUNT 29940 a 45687 c 45637 g 32240 t 2266 others.  
ORIGIN

Query Match 13.9% Score 296.6; DB 2; Length 155770;  
Best Local Similarity 82.2% Pred. No. 1e-43; Indels 83; Gaps 2;  
Matches 402; Conservative 0; Mismatches 4;

QY 1710 GCTGCGTTGACTGTGCTGAGTGACAGAACTCTGGACCCACGTGAGAGGCTGAA 1769  
|||  
Db 21540 GCAGCGTTGACTGTGCTGAGTGACAGAACTCTGGACCCACGTGAGAGGCTGAA 21481  
QY 1770 GCGGAGAGGAGGAGCCGCCG-GTGGGGAAGTGAATCTAGAGGAGGAGGAGAGGAC 1828  
|||  
Db 21480 GCGGAGAGGAGGAGCCGCCGNGTGGGACTGATCTAGGCGAGCGAGAGGAC 21421  
QY 1829 TGGACACGCTGGGGCTGGGGCTACAGGGCGGATCCCAACGGCTACTGTCTAGAC 1888  
|||  
Db 21420 TGGACACGCTGGGGCTGGGGCTACAGGGCGGATCCCAACGGCTACTGTCTAGAC 21361  
QY 1889 TCAGCGTCA----- 1898  
|||  
Db 21360 TCAGCATGCAAGTGGGGCGGGCCAGGCCAGGGCTGGGGGCAAGCTGGGGGCTGG 21301  
QY 1899 -----AGAGACCTCTGGGGGAGCGCCCTGCTC 1926  
|||  
Db 21300 AGGTGGGCGCTCTGAGTACCCCTCTCTGTAGAGGCCCTCTCGGGAGCGCCCTGCTC 21241  
QY 1927 CTAGACCTTGACCTGTTCTCACCGTCCCTGGACATGCTCTAGCCTCCACCCCTGGCTGA 1986  
|||  
Db 21240 CTAGACCTTGACCTGTTCTCACCGTCCCTGGACATGCTCTAGCCTCCACCCCTGGCTGA 21181  
QY 1987 GGGCCCCACTCTGCTGGGCCCAAGCCCTGCTGGGATCCCGGCTGGCCAGAGCAGG 2046  
|||  
Db 21180 GGGCCCCACTCTGCTGGGCCCAAGCCCTGCTGGGATCCCGGCTGGCCAGAGCAGG 21121  
QY 2047 CACGGGTATCCCGTTCACCCCAAGAGAACTGCGCTCAGTAAAGGGAACATGCCCC 2106  
|||  
Db 21120 CACGGGTATCCCGTTCACCCCAAGAGAACTGCGCTCAGTAAAGGGAACATGCCCC 21061  
QY 2107 CTGCAGACA 2115  
|||  
Db 21060 CTGCAGACA 21052

Search completed: December 7, 2001, 11:24:02  
Job time: 4939 sec





QY	1501	CCAGAGACCTGGACACGCTGTGACCCCAAGGACGCTGAGACGTCCTCTATGCCAAGGCCCGC	1500
Db	1501	ccccaggaacctggacaacgctgtgaccccaagcgacctggaaagctctctatcccaagccgc	1500
QY	1561	CTTCTCTTCCAGAACATGAACGCGGCTCCGAATACCTTCGTGAATCATCACTCTCTCGAGGT	1620
Db	1561	cttgcttctcagaacatgaaacccggtctcgaataactctcgtgaagatactcagtccttctcgtgt	1620
QY	1621	GGGGCCCCCACGAGAGATTGGAAGCGCTCAGTCACAGAAATGTACATGGACTTTGGCC	1680
Db	1621	ggggccccccacggaaagatttgaagagcgctcaatcagcaagatgtgagatgacttggcc	1680
QY	1681	ACGTTTCATGAACGTCGGGACGAGATGGTGCGTGCCTTACCTGTCGGTGGAGTGCAGAAA	1740
Db	1681	acgttctatgaactgtcggaacggtgcggtgcgtgccttgactgtgcgtgaagtgcagaaa	1740
QY	1741	CTTCTGGGACCCACAGCTGAGGAGCGCCTGAAGGCGGAGAGACGGCACCGCCGGTGGGAGC	1800
Db	1741	cttctgggaacccaacgctggaagggccctgaagagcggaagagcggaacccgcggtgcggagc	1800
QY	1801	TGGATCTTACGGGACGGGACGAGACCTGAGACGCTGAGGCTGGGCTACAGGCGGCGC	1860
Db	1801	tggatctctacggaacgagcggaacgacctggaacgctgggctgggctacagggcggc	1860
QY	1861	ATCCCCAACGGCTACACTGTGTCTAGACCTCAGCGTGTCAAGAGACCCCTCTCGGAGACGCC	1920
Db	1861	atccccaacggctacctgctgtcctagaccctcagcggtgcaagagaccctctcgggagcgcc	1920
QY	1921	TGCCCTCTAAGACCTGGACCTGTTCTCACCCTGCTGGACCTCTCTAGCTCCACCTG	1980
Db	1921	tgcctctatgaacctgtgaacctgttctccaccgtctctggaactgtctcttagcctccaccctg	1980
QY	1981	GCGTGAAGGCGCCACTCTCTTGCCTGGGCCACGCGCTGCTGGGGATCCCGCTGGCCAGG	2040
Db	1981	gcctgaaaggccccaactccctctgtctggccccagcgctctggggaatcccccgccttggccag	2040
QY	2041	AGCAGCGACGGGTGATCCCGTTCCACCCCAAGAGAACTGCGGCTCAGTAAAGGGAACA	2100
Db	2041	agcagcgacgggtgatctcccgcttccaccaccaagaagaactcgcctcagctaaacgggaaca	2100
QY	2101	TGCCCTCTGCAGACAAAAAAGAAAAAAGAAAAA 2138	
Db	2101	tgcctcttgcagacaataaaaaaagaaaaa 2138	
RESULT 2			
AA063972	ID		
AA063972	standard; cdNA; 2129 BP.		
XX	AA063972;		
XX	08-DEC-1994 (first entry)		
XX	Megakaryocyte potentiator (PKP27).		
XX	Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;		
KW	platelet; ss.		
XX	Homo sapiens.		
XX	Key		
XX	CDS		
FT	Location/Qualifiers		
FT	97..1965		
FT	/*tag= a		
FT	/*product= Meg-Pot		
FT	211..1962		
FT	/*tag= b		
FT	/*note= "claim 1; 584 amino acid product"		
FT	misc_RNA		
FT	211..954		
FT	/*tag= C		
XX	/*note= "claim 6; 248 amino acid product"		
XX	PN		
XX	W09410312-A.		

[illegible]



```

OY 659 CTTGCAACCTGCTGGGGGCTTTGGGCCAGTGGCCGAAGTGCTGTACCCGGCTG 718
Db 662 cttgcaacctgccccggggctttggggccagtcggccgaagtgtctgtacccggctgg 721
OY 719 TGAGCTGCCCCGGAGCCCTTGAGCAGACAGCAGAGAGCCAGCCAGGGCGCTTGACAG 778
Db 722 tgaagctgccccggagaccttgagccagagcaggaagcagccaggcgctctgcaggg 781
OY 779 GCGGGGAGACCCCTTAGGGCCCCCGTGCATGCTGTCTGTCCACATGTGACGCTCTGC 838
Db 782 gcggggagacccccctaacgccccccgtgacatgtctgtctccacgagtgaagctctgc 841
OY 839 GGGGCTGCTCTCCGCTGGGCCAGCCATATCCGAGCATCCGAGGAGGATCCGCG 898
Db 842 ggggctgctgctccgtgtctggccagccatcaccgagacatcccgagggatctggtg 901
OY 899 CCGCGTGGCGCAACGCTCTCTCGGAGCCATCTGGCGGACGCTGACGACCATCC 958
Db 902 ccgctgtggcgaacgctctctctcggaaccatctgtgctgagcagcctgaacggaaccatcc 961
OY 959 TCCGGCGCGGCTTCGGGAGGAGGAGAGACAGCCTGTCTTTCAGAGCAAGAGGCC 1018
Db 962 tccggcgcggtctccgcgaggagtggaagacagcctgtctctcaggaagagggccc 1021
OY 1019 GCGAGATAGAGAGAGCCTCATCTTCTTACAAAGAGTGGAGCTGGAACCTGCGTGATG 1078
Db 1022 gcgagatagagagagagctcatctctacaaagtgaggagctggaagcctgctgtggtg 1081
OY 1079 CGGCCCTGCTGGCCACCACCATGACCGCGGTGAAGCCATCCCTTACCTACGACAGC 1138
Db 1082 cggccctgtgtgacacccacagatggaacgctggaacgcatccctctacacagagcagc 1141
OY 1139 TGGACGTCTTAAGCATTAAGTGTGATGAGCTCTACCCCAAGGTTACCCGAGTCTGGA 1198
Db 1142 tggagcgtcttaagacataaactggaatgagctctaccccaagtgtaacccgagctgtgga 1201
OY 1199 TCCAGACACTGGGTACTCTTCTTCTCAAGATGAGCCCTGAGAGACATTCGCAAGTGAATG 1258
Db 1202 tccagacactgggtgactctctctcaagatgagccctgagagactctgcgaagtgaatg 1261
OY 1259 TGAGTCTCTTGAGACCTGAAAGCTTGTGTAAGTGAACAAAGGAGCAACAAATGACTC 1318
Db 1262 tgaagctctcttggaacccctggaagccttgtctggaataaaagggacacaaatgaatc 1321
OY 1319 CTCAAGCTCTCGGCGGCCCTCCACAGTGGGCACCCCTGATGACGCGCTTTGTAAGG 1378
Db 1322 ct-----cagtggtccacccctgatatgaccgcttgtgaaag 1357
OY 1379 GAAGGGGCGACCTAGACAAAGACACCTTAGACACCTGAGCCGCTTACCTGGGTACC 1438
Db 1358 gaagggcgacgtagaacaagaacacctagaacacctgacccctctacacctggtacc 1417
OY 1439 TGTGCTCTCCCTAGAGCCCGAGAGAGTGAAGTCCGTCGCCCCACAGACAGTCTGGGGGTCA 1498
Db 1418 tgtgtctctccctagagcccgagagctggaagctcgtgccccagagacatctgggggtgca 1477
OY 1499 GGCCCAAGAGACTGAGACAGTGTGACCAAGAGCAGTGAAGTGCCTCTATCCCAAGGCC 1558
Db 1478 gggcccaagagactggaacggtgagcccaagagcagctgtgaagctctctatccaaagggcc 1537
OY 1559 GCGTTCCTTCCAGAACATGAGAGGGGTCCGAATCTTGTAAGATCCAGTCTTCTGCG 1618
Db 1538 gacctgtctccagaaatgaagcgggtctgaatactctgtgaagctcagctctctcgtg 1597
OY 1619 GTGGGGCCCCCAGCAGAGATTTGAAGGCGCTCAGTCAACAGAAATGTAGACTGACTGG 1678
Db 1598 gtggggccccccagaggaatttgaagggcgtctcagcagaatgtgagcactggaactgg 1657
OY 1679 CCAGTTCATGAAGTGGGAGCGGATGGGTGCGTCCCTTGACTGTGCTGAGGAGCAGA 1738
Db 1658 ccaagctcatgaagctgcgagcggatgctgctgagcgtgagctgaggtgaggtgcaga 1717

```

```

OY 1739 AACTTGTGAGACCCACAGCTGAGGGGCTTAAGCGGAGAGACGACCGCCGCTGCGGG 1798
Db 1718 aacttctggagaccccaagtggaagggccttgaagcgggagggagcgacacgcccgtgtgg 1777
OY 1799 ACTGATCTTACGCGACGCGCAGAGACGACTGACACGCTGCGGCTGAGGCTACAGGCG 1858
Db 1778 actgatctctacgacgagcagagacgacgtggaacacgcttggtggtctgagcgtacagggcg 1837
OY 1859 GCATCCCAAGGCTACTGCTCTTAGACCTCAGCGTGCAAGAGACAGCCTTCGCGGGAGCG 1918
Db 1838 gcatcccaagcgtacctgtgtctctagacccagcgtgcaagaggcctctccgggagcgc 1897
OY 1919 CCTGCTCTTACAGACTGAGACTGTTCTTCAACGCTCTGAGCACTGCTTACCTCAACCC 1978
Db 1898 cctgctctctagagacctggaactgtctctcaacgctctggaactctctcagctcccaacc 1957
OY 1979 TGGCCTGAGGGCCCCACTCTCTTCTGCGCCACCCCTGCTGGGGATCCCGCTGGCCA 2038
Db 1958 tggcctgagggcccccaactctctgtgccccagcctgtggtggtatcccgcttgcca 2017
OY 2039 GGAGCAGGACAGGGGTGATCCCGTTCCACCCCAAGAGAACTCGCGCTCAGTAAACGGAA 2098
Db 2018 ggaagcaggaagcgggtgtatcccgcttccaccccaagagaactcgcgtcagtaaacgggaa 2077
OY 2099 CATGCCCTCTGACAGCAAAAAAAGAAAAA 2138
Db 2078 catgccctctgacagacgtaaaaaa 2117

```

```

RESULT 3
ID AAA64366 standard; cDNA; 1207 BP.
XX
AC AAA64366;
XX
DE 20-DEC-2000 (first entry)
XX
DE Soluble mesothelin related (SMR) antigen cDNA sequence.
XX
DE Mesothelin; mesothelin related antigen; MRA; malignant condition; cancer;
KW differentiation antigen; adenocarcinoma; mesothelioma; ovarian carcinoma;
KW pancreatic carcinoma; non-small cell lung carcinoma; MRA-2; SMR antigen;
XX
OS Homo sapiens.
XX
FH Key 1..1206 Location/Qualifiers
FT CDS /tag= a /product= "soluble mesothelin related (SMR) antigen"
FT /note= "no termination codon given"
PN WO200050900-A2.
XX
PD 31-AUG-2000.
XX
PE 25-FEB-2000; 2000WO-US04834.
XX
PR 26-FEB-1999; 99US-0121767.
PR 05-AUG-1999; 99US-0147404.
XX
PA (PAC1-) PACIFIC NORTHWEST RES FOUND.
PI Scholler NB, Hellstrom I, Hellstrom KE;
XX
DR WPI; 2000-572118/53.
DR P-PSDB; AAB08547.
XX
PT Diagnosing carcinomas, e.g. adenocarcinoma or ovarian carcinomas, by
PT detecting reactivity of a molecule (with an antigenic determinant and
PT present in a sample) with an antibody specific for a mesothelin related
PT antigen polypeptide
XX

```





XX Example 10; Fig 6A-B; 92pp; English.

PS The present sequence encodes a human mesothelin related antigen (MRA)-2.  
 CC Antibodies specific to MRA polypeptides are used for the detection of a  
 CC malignant condition. Mesothelin is a differentiation antigen which is  
 CC expressed in the surfaces of normal mesothelial cells and also on  
 CC certain cancer cells, including epithelial ovarian tumours and  
 CC mesotheliomas. The method is useful for detecting the presence of a  
 CC malignant condition, specifically, adenocarcinoma, mesothelioma,  
 CC ovarian carcinoma, pancreatic carcinoma or non-small cell lung carcinoma.

Sequence 985 BP; 202 A; 295 C; 318 G; 170 T; 0 other:

Query Match 40.88; Score 871.4; DB 21; Length 985;  
 Best Local Similarity 97.38; Pred. No. 1,8e-163;  
 Matches 906; Conservative 0; Mismatches 1; Indels 24; Gaps 1;

QY 970 TTCGCGGCGAGTGGAGAGACAGCTGTCCTTCCAGCAAGAAGGCCCGGAGATGAC 1029  
 Db 1 ttccggyggaagtggagagacagcctgtcctcaggaagaagccgcgagatagac 60  
 QY 1030 GAGAGCCTCATCTTCTACAGAAAGTGGAGCTGGAACCTGCTGATGCGGCCCTGCTG 1089  
 Db 61 gagagcctcatctcttaacaagaatggagctggaaagcctgctgagctgagccctgctg 120  
 QY 1090 GCCACCCAGATGGACCGCGTGAAGCCATCCCTTACACTACAGAGCACTGGAGTCTCTA 1149  
 Db 121 gccaccagatgagccgcgtgaacgcatccctcactacagagcagctgagagctccta 180  
 QY 1150 AAGCATAACTGATGATGAGCTCTACCCAAAGGTACCCTCGAGTGTGATCAGCACCTG 1209  
 Db 181 aagcataactgtagtgagctctaccacaaggttaccgccgagctgtgaltccagacctg 240  
 QY 1210 GGCTACTCTTCTCTCAAGATGAGCCCTGAGAGACTTGGCAATGTGAGCTCTCCTG 1269  
 Db 241 ggtactctctctccaagatgagccctgagacatcgcgaagtgaatgtagctcctg 300  
 QY 1270 GAGACCTGAAAGGCTTGTGAGTGAAGGACCAAGCAAAAGATCTCCAGAGCTCCT 1329  
 Db 301 gagaccctgagagctgtctgtgaagtcacaagaaggacgaatgagtcct----- 351  
 QY 1330 CGGCGGCCCTCCACAGATGCGCCACCTGATCAGCCGCTTGTGAAGGGAAGGGCGAG 1389  
 Db 352 -----caggtggccaccctgacccgtcttgaaagggaaggcgcaag 396  
 QY 1390 CTAGACAAAGACACCTTAGACACCTGACCGCTTCTTACCTGGGTACTGCTCCCTC 1449  
 Db 397 ctgagacaaagaacacctagacacctgacgcctctacacctgggtactcgtgctccctc 456  
 QY 1450 AGCCCGAGAGAGTGAAGTGTGCGTGGCCCAAGAGATCTGGGGGTCAGGCGCCAGAGAC 1509  
 Db 457 agcccgagagagctgagagctccggtgcccccaagcagcatcggggtcagggccccagac 516  
 QY 1510 CTGACACGTTGTGACCAAGGACCTGAGCTCTCTATCCCAAGGGCCGCTTGTCTTC 1569  
 Db 517 ctggacaacggtgtagcccaagagcagctgagctctctatacccaagggccgcttgcttc 576  
 QY 1570 CAGAACATGAACGGGTCCGAATACTTGTGAAGATCCAGTCTTCTGGGTGGGCCCCC 1629  
 Db 577 cagaaactgaacgggtctcgaatactctgtgaagatccagctcctccggtggggcccccc 636  
 QY 1630 ACGGAGATTTGAAGGCGCTCAGTCAAGATGAGATGAGATGAGATGAGATGAGATGAG 1689  
 Db 637 acggagatttgaagggcgctcagtcagcaagaatgtagcatgagcttgcgcaagcttcatg 696  
 QY 1690 AAGCTGGGACGAGTGGGTGCTGCTGCTTGAAGTGTGCTGAGTGTGAGAACTTCTGGGA 1749  
 Db 697 aagctgggacgagtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 756  
 QY 1750 CCCCACTGAGGAGGCTGAAGGCGGAGAGGAGCGGCGCGGCTGGGAGTGGATCCTTA 1809  
 Db 1750 CCCCACTGAGGAGGCTGAAGGCGGAGAGGAGCGGCGCGGCTGGGAGTGGATCCTTA 1809

Db 757 ccccaactgtaggggctgaaaggcgaggaagcgagccagcccgctgtaggactgataccta 816  
 QY 1810 CGGACGCGGCGAGACGACCTTGAGACGCTGGGCTGGGCTACAGGGCGGATCCCCAAC 1869  
 Db 817 cggacgcgcgagcgagcagctgagacagctgggtctgggtctacagggcgatcccccaac 876  
 QY 1870 GGCTACCTGGTCTCAGACCTCAGCGTGCAG 1900  
 Db 877 ggtcactgtgctcctagacctcagctgcaag 907

# RESULT 6

AAFP94935  
 ID AAFP94935 standard; CDNA; 396 BP.

AC AAFP94935;  
 DT 23-MAY-2001 (first entry)

DE Human ovarian cancer associated coding sequence SEQ ID NO: 126.

KW Human, ovarian cancer; vaccine; gene therapy; carcinoma; ss.

OS Homo sapiens.

PN WO200118046-A2.

PD 15-MAR-2001.

PF 08-SEP-2000; 2000MO-US24827.

PR 10-SEP-1999; 99US-0394374.

PR 03-MAY-2000; 2000US-0561778.

PR 15-AUG-2000; 2000US-0640173.

PR 07-SEP-2000; 2000US-0656668.

PA (CORI-) CORIXA CORP.

PI Xu J, Stolk JA.

DR WPI; 2001-211395/21.

XX Isolated polypeptides associated with ovarian carcinomas, and the

PT nucleic acids that encode them, useful for the prevention diagnosis and

PT treatment of ovarian cancers.

PS Claim 18; Page 157; 189pp; English.

XX The present invention provides a number of coding sequences and proteins,

CC the over-expression of which is associated with ovarian carcinoma/cancer.

CC These can be used in the diagnosis, treatment and prevention of ovarian

CC cancer, optionally by gene therapy or in the form of a vaccine. The

CC present sequence is an example of one of these sequences.

XX Sequence 396 BP; 77 A; 139 C; 111 G; 69 T; 0 other;

Query Match 10.38; Score 219.8; DB 22; Length 396;  
 Best Local Similarity 80.98; Pred. No. 9.3e-35;  
 Matches 279; Conservative 0; Mismatches 42; Indels 24; Gaps 1;

QY 1243 ATTGCAAGTGAATGTGACGTCCCTGAGACCTGGAAGCTTGTGATGAGTGAAGAA 1302  
 Db 8 actgcaagtgaaatgtgacgtccctgagacctggaagcttgcctggaagtaacaaa 67  
 QY 1303 GGGCAGGAATGACTCCTCAGGCTCTCGCGGCCCTCCACAGGTGGCCACCTGATC 1362  
 Db 68 gggcaggaatgactcctcaggtctctcgcgccctccacaggtggccacctgattc 103  
 QY 1363 GACGCGTTTGTGAAGGGAAGGGCCAGCTAGACAAAGACACCTGAGACCTGAGCGCC 1422  
 Db 104 gaccgcttctggaagggaaggcgacagctagaacaaagacacctagacacctgagccgc 163

QY 1423 TTCTACCCCTGGGACCTGTGCTCCCTACGAGCCCGAGAGAGTACCTCCGTCGCCCCAGC 1482  
|||||  
DB 164 ttctacacctgggtactcgtgtgtccctcagcccgagagagctcgtcgtcccccagc 223  
QY 1483 AGCATCTGGGCGGTGACAGGCCCGAGAGCTGACACGCTGTACCCAGGACACTGGAGCTC 1542  
|||||  
DB 224 agcatctggcggtcagcagcccgacctggacacagctggtggtctacagagcgcc 283  
QY 1543 CTCTATCCCAAGGCGCGCTTTCACAGACATGAACGGGCTCC 1587  
|||  
DB 284 atccccaagcgctacctggtctcagacctcagcatgcaagaagcc 328

## RESULT 7

AA063971

ID AA063971 standard; cDNA; 197 BP.

AC AA063971;

DT 08-DEC-1994 (first entry)

DE Meg-Pot fragment.

KW Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;

KM platelet; amplification; primer; polymerase chain reaction; PCR; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT 3..197

FT /\*tag= a

FT /note= "Meg-Pot fragment"

PN W09410312-A.

XX 11-MAY-1994.

XX 25-OCT-1993; 93WO-JP01540.

XX 23-OCT-1992; 92JP-0286153.

PR 11-NOV-1992; 92JP-0301387.

PR 09-DEC-1992; 92JP-0329546.

XX (CHUS ) CHUGAI SEIYAKU KK.

PA Hattori K, Kojima T, Oh-eda M, Yamaguchi N;

XX WPI; 1994-167467/20.

DR P-PSDB; AAR53991.

XX New megakaryocyte potentiator - for potential treatment of

PT thrombocytopenia

PS Disclosure; Page 51; 74pp; Japanese.

XX Use of the primers given in AA077809-10 in PCR resulted in the

CC fragments given in AA063971.

CC DNA encoding Meg-Pot has potential use in treatment of

CC thrombocytopenia and low platelet function.

XX Sequence 197 BP; 32 A; 66 C; 59 G; 40 T; 0 other;

SQ Query Match 9.2%; Score 197; DB 15; Length 197;

Best Local Similarity 100.0%; Pred. No. 2.7e-30;

Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 AACTCTTGGCTTCCGCTGTGCGAGGTGTCGCGCTGAGCAGGAGCGTCCGGAGC 346

DB 1 aactccttggtccgctgtgctgagaggtgtccgctgagcagagaggtgtccgagagc 60

QY 347 TGGCTGTGGCTTGGACAGAGATGTCAAGCTCTCAACAGAGAGCTGCGTGTCTGG 406

DB 61 tggctgtgaccttggcacagagaatgtcaagctctcacacagagcagctgctgtctg 120  
QY 407 CTCACGGGCTCTGTGAGGCCCCCGAGGACCTGACGCGCTCCCTGTTGAGCTGTGCTAT 466  
|||||  
DB 121 ctcaacggctctctgagcccccgagagccttgagagcctccctccattgacctgtgtctat 180  
QY 467 TCGTCAACCCAGATGCG 483  
|||||  
DB 181 tctcaaccagatgctg 197

## RESULT 8

AAH81882

ID AAH81882 standard; DNA; 255 BP.

AC AAH81882;

DT 21-SEP-2001 (first entry)

DE Rat differential transcription-associated cDNA SEQ ID 391.

KW Differential transcription; human; rat; tumour cell; cytostatic;

KM Ras modulator; Class II tumour suppressor gene; gene therapy; ss.

OS Rattus sp.

PN W0200157058-A2.

XX 09-AUG-2001.

XX 31-JAN-2001; 2001WO-EP01003.

XX 31-JAN-2000; 2000DE-1004102.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;

PI Grips M, Hellriegel M, Schmitz A, Sers C;

DR WPI; 2001-483415/52.

XX Nucleic acids differentially expressed between tumor and normal cells,

PT useful for diagnosis or therapy of tumors and for screening active

PS Claim 6; Page 488; 579pp; German.

XX This invention describes a nucleic acid (I) with differential expression

CC between tumour and normal cells and which has cytostatic activity. (I)

CC work as modulators of Ras activity by inducing expression of tumour

CC suppressor genes. (I), and polypeptides encoded by them, are useful as

CC targets for diagnosis or therapy and in screening to determine the

CC effects of an active compound (potential pharmaceutical) on a cell line,

CC particularly for diagnosis and treatment of tumors, especially by

CC modulating expression of (I) (by gene therapy, antisense RNA or ribozyme

CC methods) or by modulating the amount and/or location of (I)-encoded

CC polypeptides (by administration of the polypeptide or its activator,

CC antibody (optionally as a conjugate) or inhibitor). The method allows

CC identification of many Class II tumour suppressor genes (i.e. genes that

CC are not primary targets for tumour-initiating mutations).

CC AAH81492-AAH82376 represent the human and rat derived nucleic acid

CC fragments described in the method of the invention.

SQ Sequence 255 BP; 71 A; 71 C; 59 G; 54 T; 0 other;

Query Match 7.0%; Score 150.2; DB 22; Length 255;

Best Local Similarity 76.1%; Pred. No. 4.9e-21;

Matches 185; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1082 CCGTGTGGCCACCCAGATGAGCGCGTGAACGCCATCCCTTCACTACAGAGAGCTGG 1141

DB 2 cctgtgtgcccgcagatgacctgtgaatgaattcccttacctacgagagactca 61











[illegible]

Search completed: December 7, 2001, 10:44:45  
Job time: 2642 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2001, 10:01:57 ; Search time 55.33 seconds  
(without alignments)  
8751.310 Million cell updates/sec

Title: US-09-215-035-1

Perfect score: 2138  
Sequence: 1 AGGATTCGGCGCCGCC.....AAAAAAAAAAAAAAAAAAAA 2138

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Issued\_Patents\_NA: \*  
1: /cgn2.6/p10data/2/1na/5A.COMB.seq: \*  
2: /cgn2.6/p10data/2/1na/5B.COMB.seq: \*  
3: /cgn2.6/p10data/2/1na/6A.COMB.seq: \*  
4: /cgn2.6/p10data/2/1na/6B.COMB.seq: \*  
5: /cgn2.6/p10data/2/1na/PCTUS.COMB.seq: \*  
6: /cgn2.6/p10data/2/1na/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2138	100.0	2138	3	US-08-776-271-1 Sequence 1, Appl
2	2138	100.0	2138	3	US-09-215-035-1 Sequence 1, Appl
3	1994	93.3	2129	1	US-08-426-819A-34 Sequence 34, Appl
4	197	9.2	197	1	US-08-426-819A-32 Sequence 32, Appl
5	75	3.5	75	1	US-08-426-819A-31 Sequence 31, Appl
6	69	3.2	69	1	US-08-426-819A-30 Sequence 30, Appl
7	52.4	2.5	3407	1	US-08-253-155A-7 Sequence 7, Appl
8	46.4	2.2	2580	3	US-09-050-863-2 Sequence 2, Appl
9	46.4	2.2	5452	2	US-09-130-114-1 Sequence 1, Appl
10	46.4	2.2	9600	4	US-08-910-647-1 Sequence 1, Appl
11	46.4	2.2	10596	1	US-07-885-871-15 Sequence 15, Appl
12	46.4	2.2	10596	1	US-07-885-871-15 Sequence 15, Appl
13	46.4	2.2	10596	1	US-08-087-783A-15 Sequence 15, Appl
14	46.4	2.2	10596	1	US-08-194-088B-15 Sequence 15, Appl
15	46.4	2.2	10596	2	US-08-194-087-15 Sequence 15, Appl
16	46.4	2.2	10596	5	PCT-US93-04648-15 Sequence 15, Appl
17	46.4	2.2	71989	4	US-09-443-501A-2 Sequence 2, Appl
18	45.2	2.1	1505	1	US-07-915-246-1 Sequence 1, Appl
19	44.6	2.1	4257	2	US-08-690-473-1 Sequence 1, Appl
20	44.6	2.1	4257	4	US-09-259-821A-1 Sequence 1, Appl
21	44.6	2.1	4257	4	US-08-843-659-1 Sequence 1, Appl
22	44.6	2.1	12001	1	US-08-458-568A-11 Sequence 11, Appl
23	44	2.1	1771	2	US-08-533-669A-7 Sequence 7, Appl
24	43.2	2.0	20235	1	US-07-642-734C-3 Sequence 3, Appl
25	43.2	2.0	20235	1	US-08-439-009A-3 Sequence 3, Appl
26	43.2	2.0	20235	3	US-08-869-696-8 Sequence 8, Appl
27	43	2.0	1690	3	US-08-869-696-8 Sequence 8, Appl

28	42.6	2.0	4524	2	US-08-845-998-7 Sequence 7, Appl
29	42.6	2.0	4524	3	US-09-206-537-7 Sequence 7, Appl
30	42.6	2.0	4524	4	US-09-430-854-7 Sequence 7, Appl
31	42.4	2.0	1282	3	US-08-869-696-21 Sequence 21, Appl
32	42.4	2.0	1372	3	US-08-869-696-22 Sequence 22, Appl
33	42.4	2.0	5529	3	US-08-869-696-1 Sequence 1, Appl
34	41.8	2.0	31571	1	US-08-323-443B-1 Sequence 1, Appl
35	41.6	1.9	4411529	4	US-09-103-840A-1 Sequence 1, Appl
36	41.4	1.9	1931	2	US-05-130-114-2 Sequence 2, Appl
37	41.4	1.9	2214	3	US-08-864-038A-1 Sequence 1, Appl
38	41.4	1.9	3331	3	US-08-864-038A-2 Sequence 2, Appl
39	41.4	1.9	3331	3	US-08-864-038A-4 Sequence 4, Appl
40	41.2	1.9	1995	1	US-08-425-069-3 Sequence 1, Appl
41	41.2	1.9	1995	2	US-08-317-844B-3 Sequence 3, Appl
42	41.2	1.9	2793	1	US-08-209-747-1 Sequence 1, Appl
43	41.2	1.9	2793	1	US-08-458-298-1 Sequence 1, Appl
44	41	1.9	35100	2	US-08-770-379-17 Sequence 17, Appl
45	41	1.9	35100	4	US-08-757-669A-17 Sequence 17, Appl

#### ALIGNMENTS

RESULT 1  
US-08-776-271-1  
Sequence 1, Application US/08776271  
Patent No. 6083502  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
TITLE OF INVENTION: Mesothelin, a Differentiation Antigen  
TITLE OF INVENTION: Present on Mesothelium, Mesotheliomas and Ovarian Cancers  
TITLE OF INVENTION: and Methods and Kits for Targeting the Antigen  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/776,271  
FILING DATE: 01-DEC-1998  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/00224  
FILING DATE: 03-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/010,166  
FILING DATE: 05-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Paris, Susan K. 41,739  
REGISTRATION NUMBER: 41,739  
REFERENCE/DOCKET NUMBER: 015280-259100US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2138 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 100..1986







```

QY 1201 CAGACCTGGGCTACCTCTTCTCAAGATGAGCCCTGAGGACATTCGCAAGTGGATGTG 1260
    |||||
Db 1201 CAGACCTGGGCTACCTCTTCTCAAGATGAGCCCTGAGGACATTCGCAAGTGGATGTG 1260
QY 1261 ACCTGCTGAGAGACCTGGAAGGCTTTTCTTGAATGACAAAGGCGACGAATAGTCT 1320
    |||||
Db 1261 ACCTGCTGAGAGACCTGGAAGGCTTTTCTTGAATGACAAAGGCGACGAATAGTCT 1320
QY 1321 CAGGCTCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1380
    |||||
Db 1321 CAGGCTCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1380
QY 1381 AGGGGCGAGCTAGACAAAGACACCTTAGACACCTGACCGCTTCTACCTGAGTACTG 1440
    |||||
Db 1381 AGGGGCGAGCTAGACAAAGACACCTTAGACACCTGACCGCTTCTACCTGAGTACTG 1440
QY 1441 TGTCTCTCAAGCCCGAGAGAGCTGAGTCCGTCGCCCCCAGACAGCATCTGGGGCGTCA 1500
    |||||
Db 1441 TGTCTCTCAAGCCCGAGAGAGCTGAGTCCGTCGCCCCCAGACAGCATCTGGGGCGTCA 1500
QY 1501 CCCCAGAGACCTGGGACAGTGTGACCCCAAGGAGCTGGACGCTCTTATCCCAAGCCCGC 1560
    |||||
Db 1501 CCCCAGAGACCTGGGACAGTGTGACCCCAAGGAGCTGGACGCTCTTATCCCAAGCCCGC 1560
QY 1561 CTTGCTTTCCAGAACATGAACGGGGTCCGAATACTTCTGAGAGATCCAGTCTTCTGGGT 1620
    |||||
Db 1561 CTTGCTTTCCAGAACATGAACGGGGTCCGAATACTTCTGAGAGATCCAGTCTTCTGGGT 1620
QY 1621 GGGGCCCCCAGAGAGATTTGAGGGCGCTCAGTACAGCAATGTGAGCANTGAGCTTGGCC 1680
    |||||
Db 1621 GGGGCCCCCAGAGAGATTTGAGGGCGCTCAGTACAGCAATGTGAGCANTGAGCTTGGCC 1680
QY 1681 ACCTTCATGAAGCTGGGAGAGAGATGGTGTGCTGCTGATGTGAGTGGAGTGCAGAAA 1740
    |||||
Db 1681 ACCTTCATGAAGCTGGGAGAGAGATGGTGTGCTGCTGATGTGAGTGGAGTGCAGAAA 1740
QY 1741 CTTCTGGGACCCCACTGAGAGGGCTGAGAGCGGAGAGCGCACCGCGCTGCGGGAC 1800
    |||||
Db 1741 CTTCTGGGACCCCACTGAGAGGGCTGAGAGCGGAGAGCGCACCGCGCTGCGGGAC 1800
QY 1801 TGGATCTTACGAGCGGCGGAGAGAGCTGAGACAGCTGGGGGCTGAGGGCGGC 1860
    |||||
Db 1801 TGGATCTTACGAGCGGCGGAGAGAGCTGAGACAGCTGGGGGCTGAGGGCGGC 1860
QY 1861 ATCCCAAGGCTACCTGCTGCTGAGCTGAGCTGAGCTGCAAGAGACCTTCTCGGGAGCGCC 1920
    |||||
Db 1861 ATCCCAAGGCTACCTGCTGCTGAGCTGAGCTGAGCTGCAAGAGACCTTCTCGGGAGCGCC 1920
QY 1921 TGCCTCTAGAGACCTGAGACCTGCTCACCCTGCGCATGCTCTAGCTTCACCTG 1980
    |||||
Db 1921 TGCCTCTAGAGACCTGAGACCTGCTCACCCTGCGCATGCTCTAGCTTCACCTG 1980
QY 1981 GCTTGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2040
    |||||
Db 1981 GCTTGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2040
QY 2041 AGCAGGACGGGTATCCCGCTTCCACCCCAAGAGAACTCGCGCTAGTAAGCGGAGACA 2100
    |||||
Db 2041 AGCAGGACGGGTATCCCGCTTCCACCCCAAGAGAACTCGCGCTAGTAAGCGGAGACA 2100
QY 2101 TGGCCCTGCGAGACAAAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 2138
    |||||
Db 2101 TGGCCCTGCGAGACAAAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 2138

```

```

: APPLICANT: Hatfori, Kunihito
: TITLE OF INVENTION: Genes Coding for Megakaryocyte
: TITLE OF INVENTION: Potentiator
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: P.O. Box 747
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22040-0747
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/426,819A
: FILING DATE: 21-Apr-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Murphy Jr., Gerald M.
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 230-107P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8050
: INFORMATION FOR SEQ ID NO: 34:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2129 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: IMMEDIATE SOURCE:
: CLONE: PKP027
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 97..1965
: NAME/KEY: misc_feature
: LOCATION: 1873
: OTHER INFORMATION: /note="this residue is A in
: OTHER INFORMATION: PKP021"
: US-08-426-819A-34

```

Query Match 93.3%; Score 194; DB 1; Length 2129;  
 Best Local Similarity 97.0%; Pred. No. 0;  
 Matches 2075; Conservative 0; Mismatches 35; Indels 30; Gaps 3;

```

QY 5 ATTCGGTGGCGGCGCACTCCCTGCTGTGACGGCGGAGACAGAGACTCCGGTGGAC 64
    |||||
Db 2 AATTCGGGACGAGGCGCACTCCCTGCTGTGACGGCGGAGACAGAGACTCCGGTGGAC 61
QY 65 CCAGGCTCTCCCTCCCTGAGATCTACACAGACCATGGCTTG-CAACGGCTGAGACCC 123
    |||||
Db 62 CCAGGCTCTCTCTCTCTGAGATCTACACAGACCATGGCTTGCAACGGCTGAGACCC 121
QY 124 TGTGGTCTGTGGGAGACCGCC---TGGACCTTCCTGCTCTTCAAGCTCG 178
    |||||
Db 122 TGTGGTCTGTGGGAGACCGCCCTCGGAGAGCTCTCTCTCTTCAAGCTCG 181
QY 179 GATGGGTGATCCCGCGAGAGACCTGGGTGAGAGACAGGAGAGAGTCTCCCCCTGG 238
    |||||
Db 182 GATGGGTGATCCCGCGAGAGACCTGGGTGAGAGAGAGGAGAGAGTCTCACCCCTGG 241
QY 239 GGGGAGCTCTGACAAACCCCAATATTTCCAGAGCTTCCCTGCGCAATCTCTGGCT 298
    |||||
Db 242 ACGGAGTCTCTGCGCAACCCCAATATTTCCAGAGCTTCCCTGCGCAATCTCTGGCT 301
QY 299 TCCCGTGTGCGGAGGTGTCCGCGCTGAGCAGGAGCGGTGTCCGGAGCTGTGGCT 358

```

RESULT 3  
 US-08-426-819A-34  
 : Sequence 34, Application US/08426819A  
 : Patent No. 5723318  
 : GENERAL INFORMATION:  
 : APPLICANT: Yamaguchi, No. 5723318om1  
 : APPLICANT: Kojima, Tetsuo  
 : APPLICANT: Oh-Eda, Masayoshi

|||||  
 Db 302 TCCCGTGTGCGAGAGTGTCCGCGCTGAGCAGGAGGCTGTCCGGAGCTGGCTGTGAGCT 361  
 QY 359 TGGCAGAGAAATGTCAAGCTCTCAACAGAGAGCTGGCGCTGTGTGGCTACACGGCTCT 418  
 Db 362 TGGCAGAGAAATGTCAAGCTCTCAACAGAGAGCTGGCGCTGTGTGGCTACACGGCTCT 421  
 QY 419 CTGAGCGCCCGAGAGAGCTGTGAGCGCCCTCCATTTGACAGCTGTCTATTCTTCAACCCAG 478  
 Db 422 CTGAGCGCCCGAGAGAGCTGTGAGCGCCCTCCATTTGACAGCTGTCTATTCTTCAACCCAG 481  
 QY 479 ATGCGTTCTCGGGGCCCCAGAGCTGTGACCCGTTCTTCCGCTATCAGAGAGCCAGT 538  
 Db 482 ATGCGTTCTCGGGGCCCCAGAGCTGTGACCCGTTCTTCCGCTATCAGAGAGCCAGT 541  
 QY 539 TGGACCTGCTCCCGAGAGGGGGCTCCCGAGAGAGAGCGGCTGTGCGCGCTGTGAGCT 598  
 Db 542 TGGACCTGCTCCCGAGAGGGGGCTCCCGAGAGAGAGCGGCTGTGCGCGCTGTGAGCT 601  
 QY 599 GCTGGGGTGTGCGGGGGTCTCTGCTGAGCGAGGCTGATGTGCGGGCTCTGGAGAGCTGG 658  
 Db 602 GCTGGGGTGTGCGGGGGTCTCTGCTGAGCGAGGCTGATGTGCGGGCTCTGGAGAGCTGG 661  
 QY 659 CTTCGCACTGCTGGGCGCTTTTGTGTGCGAGTGTGCGGAGTGTGCTACCCGCTGG 718  
 Db 662 CTTCGCACTGCTGGGCGCTTTTGTGTGCGAGTGTGCGGAGTGTGCTACCCGCTGG 721  
 QY 719 TGAAGTGGCCCGGAGACCCCTGAGACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 778  
 Db 722 TGAAGTGGCCCGGAGACCCCTGAGACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 781  
 QY 779 GCGGGGAGACCCCTGAGACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 838  
 Db 782 GCGGGGAGACCCCTGAGACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841  
 QY 839 GGGGCGCTGCTGGCGCTGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898  
 Db 842 GGGGCGCTGCTGGCGCTGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901  
 QY 899 CGCGTGGGAG 958  
 Db 902 CGCGTGGGAG 961  
 QY 959 TCCGCGCGCGCTTCCGCGCGGAG 1018  
 Db 962 TCCGCGCGCGCTTCCGCGCGGAG 1021  
 QY 1019 GCGAGATAG 1078  
 Db 1022 GCGAGATAG 1081  
 QY 1079 CGGCGCGCTGCGCGCGAG 1138  
 Db 1082 CGGCGCGCTGCGCGCGAG 1141  
 QY 1139 TGGAGCTCTTAAGAGATTAAG 1198  
 Db 1142 TGGAGCTCTTAAGAGATTAAG 1201  
 QY 1199 TCCAG 1258  
 Db 1202 TCCAG 1261  
 QY 1259 TGAAGTGGCGAG 1318  
 Db 1262 TGAAGTGGCGAG 1321  
 QY 1319 CTGAGAGCTCTCGGCGGCGCTCCAGAGAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1378  
 Db 1322 CT-----CAAGTGGCGAG 1357  
 QY 1379 GAAGGGGCGAG 1438

Db 1358 GAAGGGGCGAG 1417  
 QY 1439 TGTGCTCCCTCAGCGCCCGAG 1498  
 Db 1418 TGTGCTCCCTCAGCGCCCGAG 1477  
 QY 1499 GGGCGCGAG 1558  
 Db 1478 GGGCGCGAG 1537  
 QY 1559 GCGTGTCTTCAG 1618  
 Db 1538 GCGTGTCTTCAG 1597  
 QY 1619 GTGGGGGCGCGAG 1678  
 Db 1598 GTGGGGGCGCGAG 1657  
 QY 1679 CCAAGTTCATGAAG 1738  
 Db 1658 CCAAGTTCATGAAG 1717  
 QY 1739 AACTTCTGGAG 1798  
 Db 1718 AACTTCTGGAG 1777  
 QY 1799 ACTGATCTCAG 1858  
 Db 1778 ACTGATCTCAG 1837  
 QY 1859 GCATCCCAAG 1918  
 Db 1838 GCATCCCAAG 1897  
 QY 1919 CTTGCTCTCTAG 1978  
 Db 1898 CTTGCTCTCTAG 1957  
 QY 1979 TGGCTAG 2038  
 Db 1958 TGGCTAG 2017  
 QY 2039 GAG 2098  
 Db 2018 GAG 2077  
 QY 2099 CATGCCCCCTGAG 2138  
 Db 2078 CATGCCCCCTGAG 2117

RESULT 4  
 US-08-426-819A-32  
 ; Sequence 32, Application US/08426819A  
 ; Patent No. 5723318  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yamaguchi, No. 5723318oml  
 ; APPLICANT: Kojima, Tetsuo  
 ; APPLICANT: Oh-Eda, Masayoshi  
 ; APPLICANT: Hatori, Kunihito  
 ; TITLE OF INVENTION: Genes Coding for Megakaryocyte  
 ; TITLE OF INVENTION: Potentiator  
 ; NUMBER OF SEQUENCES: 37  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
 ; STREET: P.O. Box 747  
 ; CITY: Falls Church  
 ; STATE: Virginia  
 ; COUNTRY: USA  
 ; ZIP: 22040-0747  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 197 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..197  
OTHER INFORMATION: /label=fragment  
OTHER INFORMATION: /note="amplified product from pool D using primers 3a51 and  
OTHER INFORMATION: 75a1"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..197  
US-08-426-819A-32

Query Match 9.2%; Score 197; DB 1; Length 197;  
Best Local Similarity 100.0%; Pred. No. 6,7e-34;  
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 AACTCTTGCTGCTCCGCTGCGAGGTGTCGCGCCCTGAGCAGCAGCGTGTCCGGAGC 346  
DB 1 AACTCTTGCTGCTCCGCTGCGAGGTGTCGCGCCCTGAGCAGCAGCGTGTCCGGAGC 60  
QY 347 TGGCTGTGGCTGGCAGAGATGTCATCAAGCTCTCAACGACGACCTGGCGTCTGG 406  
DB 61 TGGCTGTGGCTGGCAGAGATGTCATCAAGCTCTCAACGACGACCTGGCGTCTGG 120  
QY 407 CTCACCGGCTCTGAGACCCCGCAGGACCTGAGACGCCCTCCATTGGACCTGCTGAT 466  
DB 121 CTCACCGGCTCTGAGACCCCGCAGGACCTGAGACGCCCTCCATTGGACCTGCTGAT 180  
QY 467 TCCTCAACCCAGATGCG 483  
DB 181 TCCTCAACCCAGATGCG 197

RESULT 5  
US-08-426-819A-31  
Sequence 31, Application US/08426819A  
Patent No. 5723318  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, No. 5723318om1  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihito  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
US-08-426-819A-31

Query Match 3.5%; Score 75; DB 1; Length 75;  
Best Local Similarity 100.0%; Pred. No. 7.2e-08;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 CTCCTGAGCCCCCGAGACCTGAGACCCCTCCATTGACCTGCTGCTCAAC 474  
DB 1 CTCCTGAGCCCCCGAGACCTGAGACCCCTCCATTGACCTGCTGCTCAAC 60  
QY 475 CCAGATGGTCTCG 489  
DB 61 CCAGATGGTCTCG 75

RESULT 6  
US-08-426-819A-30  
Sequence 30, Application US/08426819A  
Patent No. 5723318  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, No. 5723318om1  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihito  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000



Matches 113; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```

OY 1632 GGAGGATTGAAGGGCGCTCAGTACAGATGTGAGCATGTGGCCAGTTTCATGAA 1691
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 969 GCAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1028
OY 1692 GCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1751
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1029 GGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1088
OY 1752 CCACGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1811
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1089 GGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1148
OY 1812 GCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1149 GGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1192

```

## RESULT 9

```

US-09-130-114-1/c
: Sequence 1, Application US/09130114
: Patent No. 5976807
: GENERAL INFORMATION:
: APPLICANT: Horlick, Robert A.
: APPLICANT: Damej, Bassam B.
: APPLICANT: Robbins, Alan K.
: TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
: TITLE OF INVENTION: From Multiple Transfected Epilomes
: FILE REFERENCE: 0867/1D903US1
: CURRENT APPLICATION NUMBER: US/09/130.114
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 5452
: TYPE: DNA
: ORGANISM: VEBNA
US-09-130-114-1

```

Query Match 2.2%; Score 46.4; DB 2; Length 5452;

Best Local Similarity 50.4%; Pred. No. 0.28; Mismatches 111; Indels 0; Gaps 0;

```

OY 1632 GGAGGATTGAAGGGCGCTCAGTACAGATGTGAGCATGTGGCCAGTTTCATGAA 1691
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1836 GCAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1777
OY 1692 GCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1751
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1776 GGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1717
OY 1752 CCACGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1811
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1716 GGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1657
OY 1812 GCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1656 GGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1613

```

## RESULT 10

```

US-08-910-647-1
: Sequence 1, Application US/08910647
: Patent No. 6251433
: GENERAL INFORMATION:
: APPLICANT: Zuckermann et al.
: TITLE OF INVENTION: Compositions and Methods for
: TITLE OF INVENTION: Polynucleotide Delivery
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation

```

STREET: 4560 Horton Street

CITY: Emeryville

STATE: California

COUNTRY: U.S.A.

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/910.647

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Fujita, Sharon M.

REGISTRATION NUMBER: 38,459

REFERENCE/DOCKET NUMBER: 1218.002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 923-2706

TELEFAX: (510) 655-3542

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 9600 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-910-647-1

Query Match 2.2%; Score 46.4; DB 4; Length 9600;

Best Local Similarity 50.4%; Pred. No. 0.32; Mismatches 113; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```

OY 1632 GGAGGATTGAAGGGCGCTCAGTACAGATGTGAGCATGTGGCCAGTTTCATGAA 1691
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1015 GCAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1074
OY 1692 GCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1751
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1075 GGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1134
OY 1752 CCACGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1811
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1135 GGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1194
OY 1812 GCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1195 GGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1238

```

## RESULT 11

```

US-07-884-811-15
: Sequence 15, Application US/07884811
: Patent No. 5316921
: GENERAL INFORMATION:
: APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
: TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: palin (Genentech)
: CURRENT APPLICATION DATA:

```







Db 2549 GCAGGACGAGGAGGGGCGAGGACGAGAGGGGCGAGAGGGGCGAGAGCA 2608  
QY 1692 GCTGCGGACGGATTGGGCTGCTGCCGTTGACTGTGCTGAGGTGCAGAACTTCTGGGACC 1751  
Db 2609 GGAGGGCGAGGACGAGGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGGGGCA 2668  
QY 1752 CCACGTGAGGGGCGCTGAAGCGGAGAGCGGCAACCGCCGCTGCGGGACTGATCTTACG 1811  
Db 2669 GGAGCAGAGAGGGGCGAGGAGGGGCGAGAGCAGAGAGGGGCGAGAGGGGCGAGAGCA 2728  
QY 1812 GCAGCGGCGAGGACGACTGACACGCTGGGGCTGGGGCTACAGG 1855  
Db 2729 GGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGAGGGGCGAGG 2772

Search completed: December 7, 2001, 11:20:41  
Job time: 4724 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2001, 09:53:47 ; Search time 2235.52 Seconds  
(without alignments)  
16761.848 Million cell updates/sec

Title: US-09-215-035-1  
Perfect score: 2138  
Sequence: 1 AGCAATCGCGTGGCCGCC.....AAAAAAAAAAAAAAAAAAAA 2138

Scoring table: IDENTITY\_NTC  
Gapop 10.0 , Gapext 1.0

Searched: 17159718 seqs, 8763200856 residues

Total number of hits satisfying chosen parameters: 34319436

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_NA\_Main:\*

```
1: /cgn2_6/ptodata/2/pna/PCRNUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US098_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US099_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099G_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099H_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099I_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099J_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US099K_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2042.6	95.5	2125	65	US-60-324-185-24511 Sequence 24511, A
2	2034.8	95.2	2126	50	US-60-172-660-22854 Sequence 22854, A
3	2034.8	95.2	2126	54	US-60-213-359-6105 Sequence 6105, A
4	1994	93.3	2129	8	US-08-426-819-34 Sequence 34, Appl
5	1981.8	92.7	2433	31	US-09-867-701-10869 Sequence 10869, A
6	1976.8	92.5	2289	25	US-09-644-668-9499 Sequence 9499, Ap
7	1976.8	92.5	2289	25	US-09-652-123-9530 Sequence 9530, Ap
8	1976.8	92.5	2289	25	US-09-652-127-9551 Sequence 9551, Ap
9	1976.8	92.5	2289	25	US-09-698-010-11861 Sequence 11861, A
10	1976.8	92.5	2289	27	US-09-699-998-8662 Sequence 8662, Ap
11	1976.8	92.5	2289	28	US-09-710-280-2725 Sequence 2725, Ap
12	1976.8	92.5	2289	28	US-09-710-281-4128 Sequence 4128, Ap
13	1976.8	92.5	2289	28	US-09-710-286-2845 Sequence 2845, Ap
14	1976.8	92.5	2289	29	US-09-726-171-1716 Sequence 1716, Ap
15	1999.6	46.8	1108	49	US-60-164-285-821 Sequence 821, App
16	992.8	46.4	1207	19	US-09-513-597-14 Sequence 14, Appl
17	983.8	46.0	1198	19	US-09-513-597-3 Sequence 3, Appl1
18	871.4	40.8	985	19	US-09-667-701-10668 Sequence 10668, A
19	812.4	38.0	881	31	US-09-667-701-10668 Sequence 10668, A
20	512	23.9	630	25	US-09-644-668-7769 Sequence 7769, Ap
21	482.2	22.6	487	31	US-09-867-701-10196 Sequence 10196, A
22	482.2	22.6	487	31	US-60-207-484-10196 Sequence 10196, A
23	478	22.4	486	31	US-09-866-555-16048 Sequence 16048, A
24	440.8	20.6	564	28	US-09-710-280-1687 Sequence 1687, Ap
25	425.2	19.9	562	17	US-09-330-360-48 Sequence 48, Appl
26	425	19.9	553	17	US-09-330-360-182 Sequence 182, App
27	421.4	19.7	544	29	US-09-726-171-762 Sequence 762, App
28	412.2	19.3	537	29	US-09-726-171-762 Sequence 762, App
29	403.2	18.9	547	29	US-09-758-458-151 Sequence 151, Appl
30	401.4	18.8	479	31	US-09-867-701-2780 Sequence 2780, Ap
31	401.4	18.8	479	31	US-60-207-484-2780 Sequence 2780, Ap
32	398.8	18.7	470	31	US-09-824-559-7925 Sequence 7925, Ap
33	398.8	18.7	470	31	US-09-824-559-7925 Sequence 7925, Ap
34	390.4	18.3	527	27	US-09-698-010-10381 Sequence 10381, A
35	382.2	17.9	455	31	US-09-867-701-4654 Sequence 4654, Ap
36	382.2	17.9	455	31	US-60-207-484-4654 Sequence 4654, Ap
37	382	17.9	425	31	US-09-867-701-1372 Sequence 1372, Ap
38	382	17.9	425	31	US-60-207-484-1372 Sequence 1372, Ap
39	375.4	17.6	460	13	US-08-671-050-3662 Sequence 3662, Ap
40	375.4	17.6	460	13	US-08-671-050-3662 Sequence 3662, Ap
41	375.4	17.6	460	13	US-08-671-050-3662 Sequence 3662, Ap
42	375.4	17.6	460	13	US-08-671-050-3662 Sequence 3662, Ap
43	375.4	17.6	460	13	US-08-671-050-3662 Sequence 3662, Ap
44					US-09-912-292-36125 Sequence 36125, A

42	370.4	17.3	441	31	US-09-867-701-2621	Sequence 2621, Ap
43	470.4	17.3	443	53	US-60-207-484-2621	Sequence 2620, Ap
44	365.2	17.1	503	27	US-09-699-998-1770	Sequence 1770, Ap
45	357.8	16.7	442	17	US-09-330-360-151	Sequence 151, App

## ALIGNMENTS

```

RESULT      1
US-60-324-185-24511
: Sequence 24511, Application US/60324185
: GENERAL INFORMATION:
: APPLICANT: Morris, MacDonald
: APPLICANT: Lal, Preeti
: APPLICANT: DiBP, DinH
: TITLE OF INVENTION: METHOD FOR THE IDENTIFICATION OF SEQUENCE POLYMORPHISMS USING
: TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCE DATABASES, AND SINGLE NUCLEOTIDE
: TITLE OF INVENTION: POLYMORPHISMS IDENTIFIED THEREBY
: FILE REFERENCE: GX-0019-1 P
: CURRENT APPLICATION NUMBER: US/60/324, 185
: CURRENT FILING DATE: 2001-09-21
: NUMBER OF SEQ ID NOS: 35862
: SOFTWARE: PERL Program
: SEQ ID NO 24511
: LENGTH: 2125
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incycle ID NO: 346520.3
US-60-324-185-24511

```

Query Match	95.5%	Score 2042.6;	DB 65;	Length 2125;
Best Local Similarity	98.3%;	Pred. No. 0;		
Matches 2086; Conservative	0;	Mismatches 29;	Indels 6;	Gaps 2;

QY	1	AGAAATTCGGGTGGCCGGCCGACATCCCTGCTGTGTATGACGGGGGAGACAGAAAGCTACCGGT	60
Db	1	aggaaattccggtctgcccggccacatccctccgtctgtgtacgcgcgagacagagactaacctgt	60
QY	61	GGACCCACAGGAGCTCCCTCCCTCCCTGGATTACACAGACCAATGGCGCTTG--CAACGGGTCGA	119
Db	61	ggaccacaagtgctctccctccctccctcgtgattcaacaagaacatgacctgtcccaacggtctcga	120
QY	120	CCCCGTGTGGTCTGTGGGGACGGCC---TGGACGCTCTGTGTCTGTCTTTACG	174
Db	121	ccccgtgtggtgtctctgtgtgagccccccgcctcgtgcagctctctgtctctcttcagc	180
QY	175	CTCGGATGGGTGTCATCCCGCGAGAGACCTTGCTGTGTAAGACAGGGAGACGGAGTGTGCCCC	234
Db	181	ctcggatgtggtgcagccctccagaaacctgtgtgagagacaggtgcaggtggtctgcgcc	240
QY	235	CTGGGGGGAGTCCGTGACACCCCCCATATACATTTCCAGCCCTTCGCCCTCCCTGGCAATCTCTT	294
Db	241	ctggaacggaagctcccgycgaacacccatacaatctccagccctccctccctccgcacaactcctc	300
QY	295	GGCTTCCCGTGTGCGGAGAGTGTCCGGGCTTAGACACGAGAGGTGTCCGGAGAGTGGCGTGTG	354
Db	301	ggtctcccggtgtgtgtaggtgtgtccgcgctgtagcaagtagctgtccggtaggcgtgctgtg	360
QY	355	GCCTTGGACAGAAAGAAATGTCATCAAGCTCTCAACAGAGAGAGCTGGCGTGTGGGCTCACCGG	414
Db	361	gcctctgcaacagaagaatgtctaaagctctcaacagtagagctgtgcgtctcgaacccg	420
QY	415	CTCTCTAGAGCCCCCGAGAGACTTGAGAGCCCTTCCATTGGAACCTGTGCTAATTCCTCAAC	474
Db	421	ctctctgagccccccgagagacctgacgcctcccatgtgacactgtctatctccaaac	480
QY	475	CCAATATGCTTCTGGGGGCCCCAGAGCTGACACCCGTTTCTTTCGCCGATACAGGAAGCC	534
Db	481	ccaaatgctgtctctgtgggcccacagacctgacccggtctctctccgcatacaacagaagcc	540

QY	535	AATGTGACCTGCTCCCGAGAGGGGGCTCCCGAGCCACACGGCTCTCCCTGCGGCTTG	594
Db	541	aatltvgaaccttgcctcccgaaagggtgtctcccgaaagcagaagctgtcctcgtcgtgtcttg	600
QY	595	GCCGTCGAGGGGTGCGCGGGGGTCTGTGCTGTAGCGAGAGCGTGAATGCTCGGGTCTGGAGGC	654
Db	601	gctctgtcgtcgtgtgtgtcgtgggtgtctctgtcgtatgagtaggtcgtatgtcgtgtctctgtgagtc	660
QY	655	CTGCGTTTCGACCTGCTCGGGCGCTTTGTGAGCCAGTGTGCGGCCGAAGTGTGTGTAACCCGG	714
Db	661	ctgtgcttgcgaacctgtcccttgggtcgtcttltgtgcgaagtcgtccgcaaggtgtcgtctaacccgg	720
QY	715	CTGGTGAGACTCCCGGGAGACCCCTGTGAGCCAGAGACACACAGAGAGGAGCCAGGGCGGCTTG	774
Db	721	ctgtgtgaactgtcctcgtggaaccccttgcagcaagagatcaggtcaggtccaggtcgtgtcttg	780
QY	775	CAGGCGCGGGAGACCCCTCTACGGGCCCCCGCTGCATGTGTCTGTCCACGATGTGAGCGCT	834
Db	781	caaggcgtggagaccctccctctacgtcccccgtctcgacaatgtctgtctccaaagtgtgaagct	840
QY	835	CTGGGGGGCCGTGCGCGCGTGTGGGGCGAGCCCATATCGAGAGATCCGAGGCGATC	894
Db	841	ctgtcgtggagctgtcgtcgtccgtgtgtgtgcgaacccaatcatccgtcagatctccgcgaaggtctc	900
QY	895	GTGGCGCGGTGGCGGCAACGCTCTCTCTCGGGAGCCATCCCTGGCGGAGAGCTGTAAAGGAGC	954
Db	901	gttgcgtcgtgtgtcgtgcgaacgtctctctctcgtgagacccaatctgtcgtcagctgtgaagcagcc	960
QY	955	ATTCCTCGCGCGCGGCTTCGCGCGGAGAGTGGAGAAACAGCCCTGTCTCTTCAAGCAAGAG	1014
Db	961	atctctctcgtcgtgtcttcgtcgtggaggtgagaaagaaagacagctgtctcttcgaagcaagaa	1020
QY	1015	GCCCGCGAGATAGCGAGACCTCATCTTCTACAAACATGTGGAGACTGTGAAGCCGTGCTG	1074
Db	1021	gtcccgtagagataagcgtgagctctcatctctcttaagaagaatgtggagctgtgaagctcgtgtg	1080
QY	1075	GATGCGGCGCTGCTGTGGCCACCCAGATGTGACCGCGGTAAAGCCATCCCTTCACTTAGAG	1134
Db	1081	gaatcgtgctcgtgtgtgtgcacacagatagtgaaacgtgtgaaagcattccctctcaactagaag	1140
QY	1135	CAGCTGAGAGCTCCCTTAAGACCTTAAACTGTGATGAGACTTACCCACAAGAGTTACCCGAGCT	1194
Db	1141	cagctgtgagatcgtccctaaagacataaacgtgagatctctaaccaagaagttaaccgcaggtct	1200
QY	1195	GTGATCCAGACACTGTGGCTACTCTTCTCTAGAGTGTAGCCCTGTAGAGCAATTGTGCAATGG	1254
Db	1201	gtgtatccgaagctcgtgtgtctacctctctcttaagaatagacccttgaagacaattctgcgaatgtg	1260
QY	1255	AATGTGAGTCCCTCGAGACCCCTGAAGGCTTCTCTGAAGTGTGACAAAGGGAGCAAAATG	1314
Db	1261	aatgtgaaagttctcctgtgagacccctgaaggtcttgtctgtgaagtaacaacaagggtcagaatagt	1320
QY	1315	AGTCTCTAGGCTCTCGAGCGGCCCTTCCCAAGTGTGGCCAGCCCTGATGTGACGCGCTTGTG	1374
Db	1321	agttctcaaggtctctcgtcgtccctctccacaaggtgtgtccacctgtatcgaacgtgtgtg	1380
QY	1375	AAGGGAAGGGCCAGCTAGACAAAGACACCCCTAGACCCCTTACCGCTTCTACCTCTGG	1434
Db	1381	aagggaagggtgtcagctatgacaagaacaccccttagacaccccttaccgtctcttacccttgg	1440
QY	1435	TACGTCGTCTCCCTACACCCCGAGAGAGTGTGCTCGTCCGCCCAAGCAAGCATCTGCGGGC	1494
Db	1441	tacctgtgtctctcccaagctcccgaaaggtcgtgagttcctcgtgtccctccagcagatcttgggtgtg	1500
QY	1495	GTCAGAGCCCCAGACACTGTGAGCTGTGACCCAAAGCCAGCTGTGAGCTCTCTATGCCAAG	1554
Db	1501	gtcaaggtcccaaggtcgtgacaggtgtgtgaccaaggtcagctgtgaagctctctatccaaag	1560
QY	1555	GCCGCGCTGTCTTTCGAGAACATGAAAGGGTCCGAACTACTTGTGTAAAGTCCAGTCTTC	1614
Db	1561	gtccgctctgtcttccagaacatgtgaacgtgtctcgaataactctgtgtaaagtccaagtctctc	1620

```

Oy 1615 CTGGGGGGGGCCCCACGAGAGATTGTGAAGGCGCTCACTACGACGAAATGACATGAC 1674
Db 1621 ctgggctggggcccccaacggagattctgaaggcctcaactgacagaatgtgagatgac 1680
Oy 1675 TTGGCCAGCTTCATCAATCACTCGACGAGATCGGtGTCGCGTTGTACTGTGTGAGAGTG 1734
Db 1681 ttggccagcttcataatgaaactcgacgagatcggtgtcgcctgtgactgtgtgagtg 1740
Oy 1735 CAGAACTTTCTGGGACCCCACTGAGAGGCGCTGAAGCGGAGAGCGGACCCGCGGTG 1794
Db 1741 cagaactctctggagcccccaactgtagggcctgaaggcgagagacggaccccggtg 1800
Oy 1795 CGGAGCTGATCTCACTGCGCAGCGGAGACGACCTGACAGCTGACGCTGGGCTAG 1854
Db 1801 cggagactgtagctccacgcaagcagcgaactgtgacacagctcgggctcgtgactaag 1860
Oy 1855 GACGCGATCCCAACGCGCTACCTGTGCTTAGACCTCAGCGTGCAGAGACCCCTTCGGGG 1914
Db 1861 ggcggtacatcccaaaaggtactcgtgtcctagaactcagcatgcaagagggcccttcggg 1920
Oy 1915 AGCGCCCTGCTCTTAGAGACGTGGAACCTGTTCMAACCGTCMGGACATGCTTAGCCGCC 1974
Db 1921 agcgctgcctctcttagagacctggaactgtcttcaacgctccgggaactgtctctagccctcc 1980
Oy 1975 ACCCTGGCCTGAGGGGCCCCACCTCCCTTGCTGCGCCACGACCTCGTGGGGATCCCCGCGTG 2034
Db 1981 accctgscctgtagggggcccaactcctctgtcgtgccccagcctgtctgggataccccgctg 2040
Oy 2035 GCCAGGACAGGACAGGCGTATCCCGTTCCACCCCAAGAGAACTCGCGCTAGTAAACG 2094
Db 2041 gccaggagcaagacaggggtgtgtcccgcttcaccccaagagaactcgctccagtaaacg 2100
Oy 2095 GGAACATGCCCCCTGACAGACA 2115
Db 2101 ggaacatgccccctgacagaca 2121

RESULT 2
US-60-172-360-22854
: Sequence 22854, Application US/60172360
: GENERAL INFORMATION:
: APPLICANT: Morris, MacDonald
: APPLICANT: Lal, Preeti
: APPLICANT: Diep, Dinh
: TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
: FILE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymorp
: CURRENT APPLICATION NUMBER: US/60/172,360
: CURRENT FILING DATE: 1998-12-16
: NUMBER OF SEQ ID NOS: 29838
: SOFTWARE: PERL Program
: SEQ ID NO 22854
: LENGTH: 2126
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID No: 346520.3
US-60-172-360-22854

```

Query Match	95.2%	Score 2034.8	DB 50	Length 2126
Best Local Similarity	98.4%	Pred. No. 0		
Matches 2088	Conservative	0	Mismatches 27	Indels 7
				Gaps 3
QY 1 AGGAATTCGGGAGCGGCCCACTCCCGTGTGTGACGCGCGGACAGAGACTACGGT	60			
DB 1 aggaatttcgggagcgcgccactcccgctgtgtgacgcgcgagagactaccggt	60			
QY 61 GGACCCAGGTCCTCCTCCTGGGATCTACACAGACCATGTGCTTG-CAACGGCTCGA	119			
DB 61 ggaccacaggctccctccctccctggatctacacagacacatgaccttcgaacagcgtcga	120			

QY	120	CCCTGTTGGTCTGTTGGGAGACGGCCC----	-TGGAGGCTCTGTCTGTCTGTTACAG	174
Db	121	ccccctgtgtggtctctgtgtggaaccccccgccttcygaagctctctgtctcttcagc	180	
QY	175	CTCGGATGAGGTCATCCCGGAGGAGACCTTGCGCTGAGAGACAGGAGCTGCTCCCC	234	
Db	181	ctcggaatggtgtgcagccctctgaaagacccctgcgtcgtgagaaacagggcagagagctgcgc	240	
QY	235	CTGGGGGAGTCTGTGACAAACCCCCATTAACATTTCCAGCTCTCCCTTCGCCCACTCTT	294	
Db	241	ctgagcaggaagctctgcgcacaacaccacataatcttcagcctctccctctcgccaactcct	300	
QY	295	GGCTTCCCGTGTGGGAGAGTGTCCGGCTGTGACACGAGACGTGTCCGGAGAGTGGCTGTG	354	
Db	301	ggctcccccgtgtgcgaggtgtctccgcgcctgtagaacagacgtgtctccggaagctgtgcgtg	360	
QY	355	GCGTTGGACAGAGATTTGTCACGTCCTCAACAGACACACTCGCGTGTGGCTACACGG	414	
Db	361	gctcttgacacagaagaatgttcaagctctcaacagagcagctgcgtctgtcttcaacccgg	420	
QY	415	CTTCTGTGACCCCCCGAGGAGACTGTGAAGCCCTCCATTGAGACCTGTGCTATTCTCTAAC	474	
Db	421	ctctcttgagcccccccgagagcctgcgaagccctcccatctgacctgcgtctatctctcaac	480	
QY	475	CCAGATGCGTTCCTGGGGCCCCAGAGCCTGTGACCCGTTTCTTCCCGCATACAGAGCC	534	
Db	481	ccagatgctgtctctgcgagcccaagcctgtgaaccgcttctctccgcatactcaagaagcc	540	
QY	535	AATGTGAGACTGTGCTCCGAGGGGGGCTCCCGAGCAGACAGGCTGTGCTGCTGGGCTGTG	594	
Db	541	aatgtgagcctgcgcaccgaaagggtgtcccgagcgcagacagcgcgtgcgtcgtgcgtcgtg	600	
QY	595	GCGTGTGAGGGTGTGTGCGGGGGTGTCTGTGCTGAGCGAGAGTGTGCGGGCTTGGGAGGC	654	
Db	601	gctcgtcgtgaggtgtgcggtgtctctgcgtgagagagctgtatgtgcggtcctctgtagagc	660	
QY	655	CTGCGTTGCGACTGCTGTGGGCGCTTTGTGTGCGGAATCGGCGAAAGTGTCTACCCGG	714	
Db	661	ctgcgtcttgagcactgcgtcctgtgcgtctgtgtgcgaatctgcgcgaagtctgtctaacccgg	720	
QY	715	CTGGTGACACTCCCGGGAGACCCCTGTGAGCAAGACACAGACAGAGGAGCGAGGGCGGCTGTG	774	
Db	721	ctggtgtgaactgcgcgggaacccctgcgacagagacagacagaggtcagcaggtgcgcgtctg	780	
QY	775	CAGGGCGGGGAGACCCCTCAAGGCGCCCGCTGCGACTGTGTCTGTCTCCACGATGAGCGCT	834	
Db	781	cagggcggggaacccccccttaagcgcgcccccgtctgacatgtctgtcttccagatgtgaagct	840	
QY	835	CTGGGGGGCTGTGCTGCGCGTGTGTGGCGACCCATATCCGACGATCCCGAGGGCATC	894	
Db	841	ctgcggtgcctgcgtgcgcgtgtgcgtgcgaagcccatatctcgcagatctccgcgaagcctc	900	
QY	895	GTGGCGCGGTGTGGGCGAAAGCTCTCTCTCGGGAGACCATCTCTGGCGGGAGCGTCAACGAGAC	954	
Db	901	gtgcgcggtgtgcgcgaacgctcctcctccgcgagaccctcttcgcgcagcgtgtgaacgagac	960	
QY	955	ATTCCTCCGCGCGCGGTTCTCCGGGGGGAAGTGGAAACACAGCGCTGTCTCTTCAAGCAAGAG	1014	
Db	961	atctcctcggccggtgtcgcgtgcggtgaagtgtgagaaagaaacgcgtctgtccttcagcagaag	1020	
QY	1015	GCCCGCGAGATGACGAGAGCTCTATCTTCTTACAGAGAGTGGAGAGCTGAGACCTGCGTG	1074	
Db	1021	gcccgcgagataagacgagagctcatctctcttacaagaagtgtggaagcctgtgcgtg	1080	
QY	1075	GATGGGCGCTGTGTGGCGCACCCAGATGAGACCGGTGAAGGCCATCCCTTACACTACAG	1134	
Db	1081	gatgtggtccctgtgcgtgcaccagatagtgacgtgtgaagcgcatactccctctcaactaag	1140	
QY	1135	CAGCTGAGCGTCTTAAAGGCAATTAAGTGTGAGACTCTTACCCCAAGATTACCCCGAGCT	1194	
Db	1141	cagctgtgagcgtctccaaagcataaactgtgattgagctcttaccacaaggttaccgccagact	1200	
QY	1195	GTGATCAAGCACTGTGGCTACTCTTCTCTCAAGTAGAGCGCTGTGAGACATTCGCAAGTGG	1254	

```
Db 1201 gtgtaccagaccctggtactctctctcaagttagcccttgagacattcgcaagtg 1260
Qy 1255 AATGTGAGCTCCCTGGAGACCTTGAAGCTTCTTGAAGTCGACAAAGGCGAGAATG 1314
Db 1261 aatgtgagctccctggagacccttaagcttcttgtaagtcacaagaagggacgaatg 1320
Qy 1315 AGTCTCAGCTCCTCGGCGCCCTCCACAGGTGCGCACCTGATGACCGCTTGTG 1374
Db 1321 agtctcagctctctctggtggtccctcccaagtggtccaccctgattgacgtcttg 1380
Qy 1375 AAGGGAAGGGGCGAGTAAAGACACCTAGACACCTGACCGCTT-CTACCGTGG 1433
Db 1381 aagggaagggcgagctagacaagacacccatagacacccgacgcgtcttgtaacctg 1440
Qy 1434 GTACTGTGCTCCCTCAAGCCCGGAGAGCTAGTCCCTGCGCCCGCCACACATCTGGCG 1493
Db 1441 gtactgtgctccctcagccccgagagctgagctcgtgccccccagcagcatctgggc 1500
Qy 1494 GGTGAGGCCCGAGGACCTGAGACGTGTGACCCAAAGGAGCTGACGCTCTATCCAA 1553
Db 1501 ggtgagggcccgagacctcgagacagctgtgacccaagagagctggaagctccctata 1560
Qy 1554 GGCCCGCTTCTTCCAGAACATGAACGGGTCCGATATCTTCTGAAGATCCAGTCTCT 1613
Db 1561 ggcccgcttcttccagacaatgaagaggttcgaatactgtgaagatccagctct 1620
Qy 1614 CCTGGTGGGGGCGCCCGACGAGGATTTGAAGCGCTCAGACAGATGAGATGAGATGA 1673
Db 1621 cctgggtggggcccccaagagagatctgaagggcgctcagctacagcaatctgaagaa 1680
Qy 1674 CTGGGCACTTTCAGTAAGCTGCGAGCGGATGCGGTGCTGCTTGAAGTGTGAGAGT 1733
Db 1681 ctgggcaacttctgagaccctcagctgagagctgtgctgctgactgtgagctgaggt 1740
Qy 1734 GCAGAACTTTGTGGAGACCCCACTGGAGGGCTTGAAGCGGAGAGCGGACCGCCGGT 1793
Db 1741 gcaaaacttctgagaccctcagctgagaggtctgaagggcgagagcgacgcgcgt 1800
Qy 1794 GCGGAGCTGATCTACGGGAGCGGAGAGAGAGCTGAGACAGCGGTGGGGCTGACA 1853
Db 1801 gcggaactgtatctcagcagcgagcaagacgacgctggaacgctggggctgacata 1860
Qy 1854 GGGGCGCATCCCAACAGGCTACTGTGCTAGACCTAGCCTGCAAGAACCTCTCTCGG 1913
Db 1861 gggcgagatccccaacgctacgtgtctctagactcaagtgcaagggccctctcg 1920
Qy 1914 GAGCGCTGCTCTAGACCTGAGACCTGTTCTACCGTCTGCGACTGCTCTAGCCTC 1973
Db 1921 gacgacctgctcttagaacctgacgttctcaacgctctgacgtcctcctagcct 1980
Qy 1974 CACCTGGGCTGAGGGGCGCCCACTGCTTGTGGGCGCCAGCGCTGCTGGGGATCCCGCT 2033
Db 1981 caacctggctgaggggccccctctctgctgagccccctgctggggatccccgct 2040
Qy 2034 GGCCAGAGAGCAGGACGGGTGATCCCGTTTCCACCCCAAGAACTGCGCTCAGTAAC 2093
Db 2041 ggcctgagagcagagcgaggtgtgtcccgcttcaccccaagagaaactcgctcagtaac 2100
Qy 2094 GGGAACTGCCCTTCGACAGCA 2115
Db 2101 gggaaactgccccctgagaca 2122
```

## RESULT 3

US-60-213-359-6105  
Sequence 6105, Application US/60213359

GENERAL INFORMATION:

APPLICANT: Morris, MacDonald

APPLICANT: Lal, Preeti

APPLICANT: Diep, Dinh

TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using

TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymor

```
FILE OF INVENTION: Identified Thereby
FILE REFERENCE: GX-0015 P
CURRENT APPLICATION NUMBER: US/60/213,359
CURRENT FILING DATE: 2000-06-21
NUMBER OF SEQ ID NOS: 7924
SOFTWARE: PERL Program
SEQ ID NO 6105
LENGTH: 2126
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
OTHER INFORMATION: Incyte ID No: 346520.3
US-60-213-359-6105
```

Query Match 95.2%; Score 2034.8; DB 54; Length 2126;  
Best Local Similarity 98.4%; Pred. No. 0;  
Matches 2088; Conservative 0; Mismatches 27; Indels 7; Gaps 3;

```
Qy 1 AGAATTCGGTGGCGCGGCGCACTCCGCTGTGCTGAGCGCGGAGAGAGCTACCGGT 60
Db 1 agaatccggtggcgcgcgccactccgctctgtgagcgcggagagactaacggt 60
Qy 61 GGACCCAGGAGTGCCTCCCTCCCTGGGATCTACACAGACCATGGCCTTG-CACGGCTGA 119
Db 61 ggaccacagtggtccctccctccctgagatctacacagacatgaccttgccaagctcga 120
Qy 120 CCCCTGTGTGCTGTGGGAGACCGCC-----TGGCACCTCTGTCTCTCTCTACG 174
Db 121 cccctgtgtggtcctgtggtgagcccccgcctcgagcctcctgtctctctcagc 180
Qy 175 CTGGGATGGGTGATCCCGCGAGAGACCTGGCTGAGAGACAGGAGACGACTTGGCCCC 234
Db 181 ctggatggtgtgacgacctcctgagaccctgtgagagaaagggcagaggtcggccc 240
Qy 235 CTGGGGGAGTCTGTGACAAACCCCATTAATTTCCAGCTCTCCCTCGCCAACTCCTT 294
Db 241 ctggagaggtctgtgacaaaccccatataatcttcacagctctccctcgcaactcct 300
Qy 295 GGCTTCCCGTGTGGGAGAGTGTCCGGCTGAGACAGAGAGCTGTCCGGAGACTGGCTTG 354
Db 301 ggtctccggtgagaggtgtcgcgacctgagcagagagctgtcggagagctgctgtg 360
Qy 355 GCCTTGGCAGAGAAATGTCAGCTCTCAACAGAGAGCTGCGCTGTCTGGCTACCGG 414
Db 361 gcttggcagagaaatgtcaagctctcaacagagagcgtcgtcgtctgctcaacgg 420
Qy 415 CTGTCTGAGCCCCCGAGAGACTGAGCGCCTCCATTGACTGTGCTATTCTCTCAAC 474
Db 421 ctctctgagcccccgagagactggaagcctcccatctgagcctcgtctctatctcctaac 480
Qy 475 CCAAGTGGCTTCTGGGGGCGCCAGGCGCTGCACCGCTTCTCTCCCGATACGAAGGCC 534
Db 481 ccagatggttctcggggcccccaagcctgcaacccgttctctcccgatcaagaagcc 540
Qy 535 AATGTGAGCTGCTCCCA-GGGGGGCTCCGAGAGAGAGCGGCTGCTGGGCGCT 593
Db 541 aatgtgagctgctcccgagaggggggtctcccgagagaaaggtctgctcgtggctct 600
Qy 594 GGCTGTGAGGGGTGTGCGGGGCTCTGTGCTGAGCGAGGCTGATGTGCGGCTTGGAGG 653
Db 601 ggctgtgaggggtgtgcggggtctctgtgagcgaggtgagtggtcgtcgagggg 660
Qy 654 CCTGGCTTGCACCTGCTGGGCGCTTGTGTGGCGGAGTGGCGGAGTGTCTTACCCG 713
Db 661 cctggcttgcacctgcttggcgcttgtgtgagcgagtgccgaagtgcgtcgtaccccg 720
Qy 714 GCTGGTGTGCTGCCCGGAGACCTGAGACAGAGAGAGAGAGAGAGAGAGAGGCGGCT 773
Db 721 gctggtgtgctgcccggagaccccttgagacagagacagagagagcagagggcgctct 780
Qy 774 GCAAGGCGGGGGAGACCCCTTACGGGCGCCCGCTGACATGCTGTCTCCAGATGAGAGC 833
```







Db 2078 CATGCCCTGCAGACACGTAATAAAAAAAAAAAAAAAAAA 2117

RESULT 5  
IIS-09-867-701-10869

```

1 Sequence 10869, Application US/09867701
2
3 GENERAL INFORMATION:
4
5 APPLICANT: Aglate, Paul A.
6
7 APPLICANT: Jones, Robert
8
9 APPLICANT: Harlocker, Susan L.
10
11 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
12
13 OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
14
15 FILE REFERENCE: 210121.497
16
17 CURRENT APPLICATION NUMBER: US/09/867,701
18
19 CURRENT FILING DATE: 2001-05-29
20
21 NUMBER OF SEQ ID NOS: 10912
22
23 SOFTWARE: FastSeq for Windows Version 4.0
24
25 SEQ ID NO 10869
26
27 LENGTH: 2433
28
29 TYPE: DNA
30
31 ORGANISM: Homo sapiens
32
33 IS-09-867-701-10869

```

Query Match	92.7%	Score 1981.8;	DB 31;	Length 2433;
Best Local Similarity	97.2%	Pred. No. 0;		
Matches 2071; Conservative	0;	Mismatches 27;	Indels 33;	Gaps 4

QY	12	TGGCGGCGCACTCCCGTCTGCTGAGCGCGGACAGAGAGCTACCGGTTGAGCCACCGGT	71
Db	330	tggcgcgcaactcccgctctgctctgtagcgcgagcaagaagctaaccggtgagaccacaggtc	389
QY	72	GCCCTCCCTCCCTGGGATCTACAGAGACCATGAGCTTG-CAAGGCTTGACCCCTGTTGGT	130
Db	390	gctccctccctctgagatctacagagaccatgagccttgccaagcgtcagaccctgtttgg	449
QY	131	CGTGTGGGGAGCCGCC-----TGGAGACCTCTGTTCTCTCTCTTACAGCTTGGAGTGGGT	185
Db	450	gtccctgtagagaccgccgcctctgagcagccctccggttctctcagctctcagctcgagatggt	509
QY	186	GCATCCCGCAGAGACCCCTGTGGTAGAGAGACGAGAGCAAGTATGCCCCCTGGGGGGAGT	245
Db	510	gcagccctcgagagacccctgctctgtagagacaag---gcagctcgacaaccccttgagcagagt	566
QY	246	CTGTACACACCCCCCATPACATTTTCAGCCTCTCCCTCGCCAACTCTTGCTGTTCCCGTG	305
Db	567	ctgtgacaaacacactaacatttccagcctctcccccctcgcaaacctcttgcttcccggtg	626
QY	306	TGGGAGAGTCTCGCGGCTTGAGCAAGGAGCGGTGTCGGGAGACTGGCTGTGGCCCTTGGACA	365
Db	627	tggcgaggtgtccgacctgagcaagagcggtgtccggagactgagctgtgaccttgagaca	686
QY	366	GAAGAAATGTAAAGCTCTCAACAGAGAGCGAGCTGTCTGCTGCTACACGCGCTCTGTGAGCC	425
Db	687	gaagaaatgtaagctctcacaacagagacgctcgctgtctgtctcagccagctctctgagcc	746
QY	426	CCCCGAGACCTTGAGCGCCTCCCATTTGAACTGTCTGTATTCTCTCAACCCAGATGCGTT	485
Db	747	ccccgagacctgtagcgccttcccatgtgacctgtctacttccctcaaccagatggtt	806
QY	486	CTGGGGGCCCAAGGCGCTGACACCGGTTCTCTCCCGATACCAAGGCCAATGTGAGCT	545
Db	807	ctcggggccccagagcctgagaccggttctctctcccgataccaagaagccaagtgtgacct	866
QY	546	GCTTCCCAAGGGGGCTCCCGAGCAGACAGCGGCTGCTCTCGGCTTGTGCTCTGCGGAGG	605
Db	867	gtctccgagggggtcctcccgagcgacaagcggctgtctcgtcggtctctggtcctgtctgggt	926
QY	606	TGTGCGGGGCTCTGTCTGTAACGACGCTGATGTGCGGGGCTCTGGAGAGCCTTGCTTCCGA	665
Db	927	tgttgcgggggtctctctctgagcagagctgtgtagtgcgggctcttgagaaagctcttgctga	986

OY	666	CTCTCCCTGGGCGGCTTTGTGGCCAGATCGGSCCAAGAGCTGTACCCCGGCTGGTAGACTG	725
Db	987	ctctccctgggcgcttcttgctgcgagtcgycgcgaatgtctgtcaaccggtctgtygaactg	1046
OY	726	CCCGGAGCCCTGGACCAAGACCAGGAGGACGACGAGCGGCGTCTGCAGGCGCGGG	785
Db	1047	cccgagacccttgagaccgaagccagcagcgaagcgaagcggtctctgcgaagcgcyg999	1106
OY	786	ACCCCCCTTAGGCCCCCGCTTCACATAGTGTCTGTCTTCACGATGACGCTCTGCGGGGCT	845
Db	1107	accccccttaggccccccctgcagcaatgctctctcaagatgacgctctgcgcyg99ct	1166
OY	846	GCTCCCGGTGCTGGGCGCAGCCCATCTATCCGCGAGCATCCCGGAGGACATCTGAGCGCGCTG	905
Db	1167	gctcccggtgctgggcgagcccatctatccgagcatctccgcagggcactgcgycgcgctg	1226
OY	906	GCGGCAACGCTCTCTCGGGACCCATCTGTGGCGGCGACGCTGTAACGGACCATCTCCGGCC	965
Db	1227	gcgcgaacgctctctctcgcgagccatctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1286
OY	966	GGGCTTCCGGCGGGAGTGGAGAAACAGCCTGTCTTTCAGCAAGAAAGCCCCGAGAT	1025
Db	1287	gggcttccggcgaggatggaagaagcaacgctctctcaagcaagaagccgcgcgaagt	1346
OY	1026	AGAGAGAGGCTCATCTCTTACAGAAAGTGGAGCTGGGAAGGCTGCGTGGATGGGGGCGCT	1085
Db	1347	agaagagagctcatctctctcaagaagtcgagcgtcggaagccgcgcgcgcgcgcgcct	1406
OY	1086	GCTCGCCACCCAGATGAGACCGGCTGAAAGCCCATCCCTTCACCTACGAGCAGCTGAGCT	1145
Db	1407	gctcgcacccagatggaacgcgcgtggaacgcacatccctctacactaagcgcgcgcgcgc	1466
OY	1146	CTTAAAGCATAAACTGATGATGACTCTTACCACACAGTTTACCCGAGCTGTGATCCACA	1205
Db	1467	cttaaaagcataaactgatagtatgactctctcaacaagaattaccccgagctctgtatctca	1526
OY	1206	CTTGGGCTACTCTTCTCTCAAGATGAGCCCTGGAGCATTTGCCAAGTGGAAATGTACATC	1265
Db	1527	cttgggcctactctctctctcaagatgagccctggagacatctgcgaagtcggaatgtaagctc	1586
OY	1266	CCTGAGAGCCCTGAAGGCTTTGCTTAAGTCGACAAAGGGCAGCAATAGTCTCTCAGGC	1325
Db	1587	cctgagagcccttgaaagccttgcttgtaagtcgaacaaagcgacggaatgagcct----	1641
OY	1326	TCTCTGGCGGCCCTCCACAGAGTGGCCACCTGATCGACCGCTTTGTGAAGGCAAGGG	1385
Db	1642	-----cagctggccacccctgacatcgacgcgctctgtgaaaggaaggg	1682
OY	1386	CCAGCTTAGAACAAGACACCCCTTAGACACCCGTGACCCGCTTCTACCTGGGTACTGTGTC	1445
Db	1683	ccagcttagacaagaacaccccttagacacccctgcgcgcctctaccccttggtacctgtgctc	1742
OY	1446	CCCTAGCCCCGAGAGAGTGAAGTCCGTGCGCCCCACACACATCTGGGCGGTACAGGCCCA	1505
Db	1743	ccctagccccgagagatgagctcgcgtgcgtgcgcccaagacatcgcgcgtctcaagcccca	1802
OY	1506	GGACCTGGACACGTTGTACCCCAAGGAGCAGCTGACGTCCTATTCCCAAGGCCCGGCTTGC	1565
Db	1803	ggacctggaacgcgtgtaaccacaagcgacgtcgcgcctctatctccaaagccccgcctgc	1862
OY	1566	TTTCCAGACATGAACGGGTCCCAATACTTCTGTGAAGATCCAGTCCCTCTGCGTGGGGC	1625
Db	1863	tttccagacaatgaaacgcgtgcgaataacttcgtgaagatccagctctctctcgtgtyg99c	1922
OY	1626	CCCCAGGAGATTTGAAGGCGCTCATGTCAGCAGATGAGACATGAGACTTGGCCACGTT	1685
Db	1923	ccccaggaagatcttgaaagcgctcagtcagcagaaatgtagcattggaacttgcgaagct	1982
OY	1686	CATTAAGCTGGGACGAGATCGGCTGCTGGCGTTGATGTGGCTGAGGTGACAGAACTCTT	1745
Db	1983	catgaagctgcgagcaggaatgcggtgtcgcgtgtgactgtgcgtgcgcgcgcgcgcgcgc	2042
OY	1746	GGGAGCCCCACGCTGAGGGGCTTGAAGCGAGAGCGGACCGCCCGGTGCGGACTGGAT	1805

|||||  
Db 2043 gggagcccaacgtgagagccctgaagcgcgagagcgagcaccgcccggctgcyggactgagt 2102  
QY 1806 CCTACGGAGGGGAGAGAGACGACCTGAGACCTGGGGCTGGGGCTACAGGGCGGATCCC 1865  
Db 2103 cctcaggaagcgagagcgacccctgagacgctgagcgctgagcgctgagcgagctatccc 2162  
QY 1866 CAACGGCTACTGGTCTCTAGACCTCAGCGGTGCAAGAGACCTCTCGGGAGCGCCCTGGCT 1925  
Db 2163 caagcgctacctgtgtccagagccctcaggtgcaagagcgccctcccgagagacccctgct 2222  
QY 1926 CCTAGGACCTGGACCTGTTCTTCACCGTCTTGACACTGCTCTAGCCTTCACCTTGCGCTG 1985  
Db 2223 cctagagactgagactgtctccacgcgtctcgtgcaactgtcctcctagcctccacccctg 2282  
QY 1986 AGGGCCCCACTTCCTGCTGGCCCCAGCCCTGCTGGGATATCCCGCTGGCGGACGAGACAG 2045  
Db 2283 agggcccaacccctctgctgccccagccctgctgagagctcccgccctgagagagag 2342  
QY 2046 GCACGGGTGATCCCGTTCACCCCAAGAGAACTCGCGCTCAGTAAACGGAGACATGCC 2105  
Db 2343 gcaagcggtgatacccgcttccaccccaagagagactcgctcagtaaacggagacatgccc 2402  
QY 2106 CCTGCAGACAAAAA 2136  
Db 2403 cctgcagacacgtataaaaaa 2433

RESULT 6  
US-09-644-868-9499  
Sequence 9499, Application US/09644868  
GENERAL INFORMATION:  
APPLICANT: White, David  
APPLICANT: Holtzman, Douglas A.  
APPLICANT: Pan, Yang  
TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
FILE REFERENCE: 1600.1168-001  
CURRENT APPLICATION NUMBER: US/09/644, 868  
CURRENT FILING DATE: 2000-08-28  
PRIOR APPLICATION NUMBER: 60/151,063  
PRIOR FILING DATE: 1999-08-27  
NUMBER OF SEQ ID NOS: 10075  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9499  
LENGTH: 2289  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-644-868-9499

Query Match 92.5%; Score 1976.8; DB 25; Length 2289;  
Best Local Similarity 97.3%; Pred. No. 0;  
Matches 2053; Conservative 0; Mismatches 27; Indels 30; Gaps 3;

QY 12 TGCGCGGACCTCCGCTCTGCTGTGACGCGGACAGAGACCTACCGGTGAGCCACCGT 71  
Db 202 tggcgcgcacactccgctctgtgtgacgcgcgcagagagactacagctgagcccaagct 261  
QY 72 GCGTCCTCCCTGGGATCTACAGACCATGCGCTTG-CAACGGCTGACCCCGTGGT 130  
Db 262 gcccctccctccctggatctacagacacalggccttgccaagctcgaccccgcttggg 321  
QY 131 CTTGTGGGACCGCC-----TGCGACGCTCTGTTCTCTGCTTTCAAGCTCGGATGGGT 185  
Db 322 gtccctgtyggagaccgccgcctcgagcctctctgtctctcagcctcgatgggt 381  
QY 186 GCATCCCGCAGAGACCTGCTGTGAGAGACGAGGAGAGTCTGCCCCCTGGGGAGT 245  
Db 382 gcaagcctctgagagcctgtgtgagagacagggcaggaagctcgccccctgagcggagt 441  
QY 246 CTTGACAAACCCCATATTTCAGGCTCTCCCTGCGCAACTCCTTGCTTCCGCTG 305

Db 442 cctgccaaccacactaacattccagcctctccctcgcgccaactccttgctcccgct 501  
QY 306 TGGGAGGTGTCCGGCTTGAGACGAGACCTGTCCGGGAGCTGTGGCTTGAGGCA 365  
Db 502 tgcggaagtgctccggccctgagacaggaagctgcccgaagctgagcttgagcga 561  
QY 366 GAAAGATGTCAAGCTCTTCACAGAGACAGCTGCTGTGCTCAGCGGCTCTCTGAGCC 425  
Db 562 gaagaatgtcaagctctcaacaagagcagctgctgtctgtctcaacggctctctggcc 621  
QY 426 CCCCAGAGACCTGAGACGCTTCCTCATTTGAGACCTGCTGCTATTCCTCAACCGATGCGTT 485  
Db 622 ccccgagacacttgagacgccccccatctgacctgtgtatctcctcaaccagatgcgt 681  
QY 486 CTGCGGGCCCGAGGCTGACACCGCTTCTCTCCGCTATCAGAGGCAATGTGGACCT 545  
Db 682 ctgcgggcccagagcctgacacagcttctctccgcacacagagacaaatgtagact 741  
QY 546 GCTCCCGAGGGGGCTCCCGAGCAGACACCGGCTGCTGCTGCGGCTGCGCTGCGG 605  
Db 742 gctcccgagggggtctcccgagcgacagcgctgctgctcgcgctcgctgctggg 801  
QY 606 TGTGCGGGGCTCTGCTGACGAGGCTGATGTCGGGCTTGAGAGCCTGCTTGCGA 665  
Db 802 tgtcgagggtctctgtcgtgagcgagctgagtgagggctctgagagcctggtctgcga 861  
QY 666 CCTGCTGTGGGCTTTGTGTGGCGAGTGGCGGAGTGTCTACCCGGCTGGTGTGAGCTG 725  
Db 862 cctgcctggggccttctgtgcccagagctcgccgaagtgctgtctacccggctggtgagctg 921  
QY 726 CCGGGGACCCCTGAGACGAGACGACGAGGAGGAGCCAGGGGCTCTGACGGCGGGGG 785  
Db 922 cccggagcccttgagacagagacagcgaggaagcgagcgagcgctcgtcagagcgagg 981  
QY 786 ACCCCCTAGGGCCCCCGCTGACATGCTGTCTCCACGATGAGACCTCTGCGGGGCT 845  
Db 982 aacccttaacggccccccgctgacatgltcgtctcaagatgagcgtctcgagggct 1041  
QY 846 GCTCCCGTGTGGGCGAGCCCATATCCGAGCATCCCGAGGCGCATCTGCGGGGCTG 905  
Db 1042 gctgcgcgtgtgtggtccagcccatcatccgagatcccgagagcatcgtggtcggtg 1101  
QY 906 GCGGCAACGCTCTCTGTGGGACCATCTGCGGAGCTGAGACGACCATCTCTCGGCC 965  
Db 1102 gcggcaacgctctctcgtgagcccatcctgagcgagcgtlgaagagacatcctccggcc 1161  
QY 966 GCGGTTCCGGGGGAGGAGGAGAGACGCTGTCTTCACGAAAGGCGCGCGAGAT 1025  
Db 1162 gcggtlcccgaggaagatggaagagacagcctgtctctcaggaagaagagcccgagat 1221  
QY 1026 AGACGAGAGCCTCATCTTCTCAAGAGAGTGGAGCTGGAACCTGCGTGTGGGCGCT 1085  
Db 1222 agaagagagctcatctctctcaagaagtgtagcctggaagccgctgtagtggtgcct 1281  
QY 1086 GCTGGCAACCATGATGGAACGCGGTGAACGCCATCCCTTACCTACGACGAGCTTGACGT 1145  
Db 1282 gctgccaacccaagatgagccgtgaaacgcctatccctcactcaagagcagctggagct 1341  
QY 1146 CTTAAGCATTAACCTGGATGGGCTTACCCCAAGGTTACCCGAGCTGTGATCGACGA 1205  
Db 1342 cctaaagcataaactgagatgagctctacccaagatgtaaccgaagcttgagatccagga 1401  
QY 1206 CTTGGGCTACTCTTCTCAAGATGAGCCCTGAGGACATTCGCAAGTGTGATGTGACGTC 1265  
Db 1402 ccttggtactctctctctcaagatgagccctggaagatctcgcaagtggaatgtagagctc 1461  
QY 1266 CTTGAGAACCTTGAAGCTTTGTTGAAGTGCACAAAGGCGACGAATGATGCTCAGGC 1325  
Db 1462 cctgagagaccttgaagccttcttgagatgtaacaagaaggagcaaatgtagctct----- 1516  
QY 1326 TCTCTGGCGGGCCCTCCACAGGTGGCGACCTCTGATGACGCGCTTGTGAAAGGAAAGGG 1385  
Db 1517 -----cagtgagcacccttgatcgacgcgttctgtgaaaggaaagg 1557

QY	1386	CCACTGTGACAAAGACACCCCTATACACACCTGACACCGCTTTCACCTGGGTACCTGTGTC	1445
Db	1558	ccagctatgacaaagaacacccctatgacacccctgacacgcctcttcaacccttggtaaccgtgtctc	1617
QY	1446	CCTGACGCCCGAGAGACTGAGCTCCCTGTCCGCCCCAGCAGACATCTTGGGCGGTCCAGGCCCA	1505
Db	1618	cctcagcccccgagagactgagctggtctcgtgtgcccccaagcatcttgggtgtcaggtcccca	1677
QY	1506	GGACCTTGACACGTGTGACCCAAAGGACAGTGGACGTCTCTATTCCAAAGGCCCGCTTGC	1565
Db	1678	ggaccttggacagctgtgacccaaaggacgttggacgtctctctatcccaaggccgccttgc	1737
QY	1566	TTTTCCAAACATGACGCGGTGCCAATACCTTCGTAAGAAATCCAGTCCCTTCGCGGGGGGCGC	1625
Db	1738	tttcagaaacaatgaaacggtgtccgaataactctgtgaaagatccacgtctctccctgggttgggc	1797
QY	1626	CCCCACGGAGGAAATTTGAAGCGCTCAGTCAGGACGAGATGTGACATGACTTGGCCAGATT	1685
Db	1798	ccccacggagaaatttgaaggcgtctcagtcagaaagatgtgagcatgtgacttggccaagtt	1857
QY	1686	CATGACACTCGGACGAGATGCGGTGCTGCCGTTTGACTGTGTGGCTGAGGTGCGAANAATTCT	1745
Db	1858	catgaaactctggagacggaatgtggtgtcgtccgttgaactgtgactgtgagctgagtcagaactctc	1917
QY	1746	GGGACCCGACGTGAGAGGAGCTTGAAGGCGGAGAGAGCGGACACGCGCGGTGCGGGACTGGAT	1805
Db	1918	gggagccccaactgtggagggcctctgaaggcggagagacgtgcacacgcccgtgtggtgagctgcat	1977
QY	1806	CTTACGGCAGCGGACGAGACGACTGTGACACGCTGGGGCTGGGGCTTACAGGGCGGATCCC	1865
Db	1978	cttactggcagcggagaaagacgaccttggacacgcttgggtgttgggtctacagggcggcatccc	2037
QY	1866	CAACGGTACTCTGTCTCTATACCTTCAGCGTGCMAAGAGACCTCTGGGGAGGCCCTGGCT	1925
Db	2038	caaacgttactactgttccctagacactcagcatlgaaagagggccctctcgggagacgacctgtcct	2097
QY	1926	CCTTAGGACCTGAGACCTGTTCTTCACCCCTCGTGACAGTGCCTCTAGCGCTCCACCTGGCCTG	1985
Db	2098	ccttaggcctcgtgagacactgttctcaacgcttctgtgacgtctcctcagctctcaacctgtgctgt	2157
QY	1986	AGGGCCCCACTCCCTTGTCTGTGGCCCCAGCCCTGCTGGGGATCCCGCGCTGCGCCAGAGCAG	2045
Db	2158	agggccccactcccttctgtgccccagccctgtcttgggatacccccgcctggccagagagag	2217
QY	2046	GCACGGGTGATCCCGTTCACCCCAAGAGACGTGCGCTCATGTAAGGGGAACATGCGC	2105
Db	2218	gcacgggtgtgttccccgttccaccccaagaagaactcgtgcctcagtaaacggaacatgtccc	2277
QY	2106	CCTGCAGACA 2115	
Db	2278	cctgcagaca 2287	
RESULT 7			
US-09-652-123-9530			
: Sequence 9530, Application US/09652123			
: GENERAL INFORMATION:			
: APPLICANT: Holtzman, Douglas A.			
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES			
: TITLE OF INVENTION: THEREFOR			
: FILE REFERENCE: 1600.1186-001			
: CURRENT APPLICATION NUMBER: US/09/652,123			
: CURRENT FILING DATE: 2000-08-30			
: PRIOR APPLICATION NUMBER: 60/151,135			
: PRIOR FILING DATE: 1999-08-30			
: NUMBER OF SEQ ID NOS: 9796			
: SOFTWARE: FASTSeq for Windows Version 4.0			
: SEQ ID NO 9530			
: LENGTH: 2289			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
US-09-652-123-9530			

Query Match		92.5%	Score 1976.8	DB 25	Length 2289
Best Local Similarity		97.3%	Pred. No. 0		
Matches 2053	Conservative	0	Mismatches	27	Indels
				30	Gaps
OY	12	TGGCCGGCCACTCCCGTGTCTGTGTAGCGCGCGAGAGAGAGCTACCGGTGAGCCAGCGT	71		
DB	202	TGGCGCGCACTCCCGTCTGTGTGTAGCGCGCGAGAGAGAGCTACCGGTGAGCCAGCGT	261		
OY	72	GGCTTCCCTTCCCTGGGATTTACACAGACATGGCTTTG--CAAGGCTTCGACCCCTGTGGT	130		
DB	262	ggcttcccttccctgggatatcacagacatgaccttgccaaagctctgcacccctgtctgg	321		
OY	131	CTGTGTGGGAGACGCC-----TGCGAGGCTCTCTGTTCCTGCTCTTCACAGCTTGATGGGT	185		
DB	322	gtctgtgtggagcccccgccctctgcagactctctgttctcgtctcttcagctctgatatggt	381		
OY	186	GCATCCCGCGAGAGACCTTGCTGTGTGAGACAGAGAGAGAGTGTGCCCCCTTGCGGGAGT	245		
DB	382	gcagaccttgagagacccttgctgagagagacagagagagctgagcccccttgacgagagt	441		
OY	246	CTGTGACACCCCCCAATTAATTTCCAGGCTTCGCCCTGCGCAACCTCTTGCGGTCCCGTG	305		
DB	442	ctgtgcacaccccaactaaatlttccagctctcccccctgcacactcttgtcttccgctg	501		
OY	306	TGCGGAGGTGTCTCGGCTCTGAGCAGCAGAGCGGTGTCCGGAGCTGTGGTGTGGCTTGACACA	365		
DB	502	tgcgagagtgctccgagctcagcagagagtggtlctcgagagcttgctgtgaccttgacaca	561		
OY	366	GAAGAATGTCAAGCTCTTCACAGAGACAGCTGGCGTGTCTGGCTTCACGCCGCTCTTGAGGC	425		
DB	562	gaagaatgtcgaagctctcaacagagacagctgctgtctgtctcaccggtctcttgaagc	621		
OY	426	CCCCAGAGACCTGTGAGAGCGCTCCACTTGTGAGCTCGTGTCTTCTTCACACCAGATGCTTT	485		
DB	622	ccccgagagacctgagagctctccatctgacctgtctgtctatctctcaaccagatgtcgct	681		
OY	486	CTTGGGAGCCCAAGGCGCTGTGACCCGTTTCTTCTTCCGCAATCAGAGGCCAATGTGACCT	545		
DB	682	ctctggggccccagagcttgacaccgttcttctctccgcagatcaagaaaggccaatgtgagact	741		
OY	546	GTTCGCCAGAGGGGGGTCCCGAGCAGACAGAGGGGTGCGTCCCTGGGGCTCTGCGCTGGGG	605		
DB	742	gtctcccgaggggggtctcccgagagagagaggtgtgtctgtcgtggtctgtgctgtggtgg	801		
OY	606	TGTGCGGGGCTCTCTGTGTGAGCAGCAGCTGTGCGGCTCTTGAGAGGCTTGAGTTCGA	665		
DB	802	tgttgcggggtgtctcttgcttgagcagagcttgaigtgctgggtcttgagagggcctgtgca	861		
OY	666	CCTGCTGTGGGCGCTTTGTGTGGCCGAGTCTGGCGCAAGTGTGTCTACCCGGGCTGTGAGCTG	725		
DB	862	cttgccttggtgctcttgcttgagcagatctggtccgaatgtctgtctaccccgctgtgtgagctg	921		
OY	726	CCCGGAGACCTTGAGACAGAGAGCAGAGAGGAGCAGCAGGGCGGCTCTCAGAGGCGGGG	785		
DB	922	cccgagagacccctcgagacagagacagcagagagtcagcagaggtgtcttgcaagggtgggg	981		
OY	786	ACCCCCCTTAGGCCCCCGCTGTGACATGTTCTGTCTTCACGATGTGAGACGCTCTGCGGGCT	845		
DB	982	accccccttaggcccccgctgaaatgtgtctgtcttccacgatagtgaagctctgtcggggct	1041		
OY	846	GCTTCCCGTGTGTGGGCAAGCCATCATCTCGCAGCATCCCGGAGAGGAGATGTGGCCGCGTG	905		
DB	1042	gtctgcgcgtgtgctgggacagccatcatctcgagcagatcccgcaagggtgaatcgtgtgcgcgtg	1101		
OY	906	GGGGAACGCTCTCTTCGGGACCAATCTCTGTGGGAGAGCTTGAAAGGAGCCATCTCTCGGCG	965		
DB	1102	ggggcaagctctctctcgtgagccatctctgtggtgagagcttgaaagacatctctcgggc	1161		
OY	966	GGGTTTCCGGCGGAAGTGTGAGACAGACAGCTGTCTTTAGGCAAGAGGCCCGCGAGAT	1025		
DB	1162	ggagttcttcggtcgagagtgtagaagaacagccgtctcttcaaggccaagaaagccgcgtgagat	1221		

```
QY 1026 AGACGAGACCTCATCTTCTCAAGAAAGTGGAGCTGGAGACCTGCGTGGAGCGCCCT 1085
    |||||||
Db 1222 agagagagagctctctctctcaagaagtggagctggaagccctgctgctgctgctcct 1281
QY 1086 GCTGGCCACCCAGATGAGACCGCGTGAACGCCATCCCTTACCTACGACGAGCTGGACGT 1145
    |||||||
Db 1282 gctggccaccagatgagacccgctggaacgcacatccctctcaacgagcagctggagct 1341
QY 1146 CCTAAGATTAACCTGATGATGCTCTACCCCAAGGTTACCCCGAGTCTGTGATTCACGA 1205
    |||||||
Db 1342 cctaagataaactgctgagctctcaacccaaagctaacccgagctctgctgctcagca 1401
QY 1206 CCTGGGCTACCTCTTCTCAAGATGAGCCCTGAGACATTGTGCAAGTGGATGTGACCTC 1265
    |||||||
Db 1402 cctgggctactctctctctcaagaatgagccctgaggaactctgcaagatggaatgagctc 1461
QY 1266 CCTGGAGACCTTGAAAGCTTGTGTAAGTGCACAAAGGCGACGAATGATCTCTCAGGC 1325
    |||||||
Db 1462 cctggagacccctgagagcttctgcttgtaaacaaagggcagcaaatgagctcct---- 1516
QY 1326 TCTCTGGGGGCGCCCTCCACAGGTGGCCACCTGATGACCGGCTTGTGAAGGGAGGGG 1385
    |||||||
Db 1517 -----cagctggccaccctgctgacccgctctgtaagggaaagg 1557
QY 1386 CCAGCTAGACAAAGACACCCCTAGACACCCCTGACCGCTTACCTGGGTACTGTGCTC 1445
    |||||||
Db 1558 ccagctagacaagaacacccctagacacccctgacccgctctctcaacccctgctgctgctc 1617
QY 1446 CCTAGACCCCGAGAGCTGAGCTCCGTCGCCCCAGCAGCATCTGGCGGTCTCAGGCCCA 1505
    |||||||
Db 1618 cctaagccccgagagctgagctctgctgccccagcagcactctgggctgcaagccccca 1677
QY 1506 GGACCTGGACACGTGTGACCCCAAGGCGAGCTGGAGTCCCTATCCCAAGGCGCGCTTGC 1565
    |||||||
Db 1678 ggaactggacacgctgctgacccaagccagctgagctctctcaacccagccgctctgctg 1737
QY 1566 TTTCACAGAACTGAACGGGCTCCGAATCTTCTGTAAGATCCAGTCTTCTCTGGGTGGG 1625
    |||||||
Db 1738 ttctcagaacatgaaagctgctgaaactctctgtaagatccagctctctcctggctgg 1797
QY 1626 CCCACGAGAGATTTGAAGGGGCTCAGTCAACAGCAATTTGACATGACATGACCTGGCACCT 1685
    |||||||
Db 1798 ccccaaggaatcttgaagggctcagtcagcaagaatgagcaatgagcttggccacgtt 1857
QY 1686 CATGAAGCTGGAGAGGATGGGTGCTGCTGCTTACTGTGAGTGAAGTGAACAATCTCT 1745
    |||||||
Db 1858 catgaagctggagacgagctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1917
QY 1746 GGGACCCACACTGGAGGGCTTGAAGGGCGAGAGGAGCGACCGCCGGTGGCGGACTGAT 1805
    |||||||
Db 1918 ggggaccccaagctggaagggctggaagggcagggagggcagccggctggcgagctg 1977
QY 1806 CCTACGGGAGGGGAGGAGCGACTGAGACAGCTGGGGCTGGGGCTACAGGGCGGATGCC 1865
    |||||||
Db 1978 cctcggcgaagggaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2037
QY 1866 CAACGGCTACTGCTCTAGACTCAGCGTCAAGAGACCTCTCGGGGAGCGCCCTGCT 1925
    |||||||
Db 2038 caaagctactgctctctagaccctcagcaatgcaagaggccctctcggggagagccctg 2097
QY 1926 CCTAGAGCTTGAGCTGTTCTCACCGTCTGGCACTGCTCTTACCTTCACCTGGCGCTG 1985
    |||||||
Db 2098 cctagagactgctgctctctagaccctcagcaatgcaagaggccctctcggggagagcc 2157
QY 1986 AGGCGCCACCTCTTGTGCTGGCGCCAGCCCTGCTGGGATCCCGCCTGGCGAGAGCAG 2045
    |||||||
Db 2158 agggcccaactcctctgctggcccaagccctgctgggagatcccgccctggccagag 2217
QY 2046 GCAGGGGATCCCGTTCACCGCCCAAGAGAACTCGCGCTAGTAAACGGGAGACATGCC 2105
    |||||||
Db 2218 gcacgggtgctcccgcttccaccccaagagaaactcgcctcagtaaaaggaacatgccc 2277
```

```
QY 2106 CCTGCAGACA 2115
    |||||||
Db 2278 cctgcagaca 2287

RESULT 8
US-09-652-127-9551
; Sequence 9551, Application US/09652127
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600.1183-001
; CURRENT APPLICATION NUMBER: US/09/652,127
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/151,134
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 10475
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9551
; LENGTH: 2289
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-652-127-9551

Query Match 92.5%; Score 1976.8; DB 25; Length 2289;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 2051; Conservative 0; Mismatches 27; Indels 30; Gaps 3;

QY 12 TGCGCGGCGCATCCGCTGCTGTGTGACCGCGGACAGAGACTACCGGTGAGCCACGGT 71
    |||||||
Db 202 tggcgcgacactcccgctgctgtgtgagcgcgcgagacagagactacgcgtggaaccacgct 261
QY 72 GCGTCCTCCCTGGGATCTACACAGACATGACCTTG-CAACGGCTGAGCCCGTGTGT 130
    |||||||
Db 262 gctcctccctgggatactacacagaccaaagccttgccaagctcgaaccccgcttggg 321
QY 131 CCTGTGGGAGACCGCC-----TGCGAGCTCTCTGCTCTTCAAGCCTCGATGGT 185
    || || ||
Db 322 gtctctgtggagcccccgcctccctcgcgcgcctctctctctcagcctcggatgg 381
QY 186 GCATCCGCGCGAGACCTGCTGTGAGAGACAGGACGAGTCTCCCCCTGGGGGAGT 245
    || || ||
Db 382 gcaacccctcgaagacccctgctggaagagacagggcaggaagcttgcgcctcgaag 441
QY 246 CCTGACAAACCCCATTAACATTTCCAGCTCTCCCTCCGCAACTCTGCTGCTCCCGT 305
    || || || || ||
Db 442 cctggccaacccaactaactatctccagcctctccctcgcgaactcttggctccgctg 501
QY 306 TGGGAGGTGTCGGGCTGAGACAGGAGCCUATGTCGGGAGCTGGCTGTGGCTTGAGACA 365
    |||||||
Db 502 tgcggaggtgtcggcgcctgagcacggagcgtgtccggagagctgtgtgcttggaca 561
QY 366 GAAGAATGTCAAGCTTTCACAGAGACACTGCGTGTGTGCTACCGGCTCTGTGAGCC 425
    |||||||
Db 562 gaagaatgtcaagcttcaacagagcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 621
QY 426 CCCGAGGAGACTTGAGAGGCGCCCATGAGACTGCTGTGTATTTCTCAACCCAGATGGGTT 485
    |||||||
Db 622 ccccgaggaacttggagcgcctcccatcggacctgtgtatctctcaaccagaatgctgt 681
QY 486 CTCGGGGCCCGAGGCTTGACCCGTTTCTTCTCCGATCAAGAGCCATATGAGACT 545
    |||||||
Db 682 ctggggcccgagcctggaacccgttctctctccgcatacgaaggccaatgtggacct 741
QY 546 GCTCCGAGGGGGCTCCCGAGCAGACAGCGGCTGCTGCTGCGGCTCTGGCGTGGG 605
    |||||||
Db 742 gctcccgaggggctcccgagcgacagcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 801
QY 606 TGTGCGGGGCTCTCTGCTGAGCGAGGCTGATGTCCGGGCTCTGGGAGGCTTGCTGCGA 665
    |||||||
Db 802 tgtcgggggtgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 861
```

QY	666	CCGCCTGGGGGGCTTTTGTGGCGAGTCGGCGCAAGTGGCTGACACCCCGGCTGGTGAAGCTG	725
Dp	862	CTTGTCTGGGAGCGCTCTTGTGGCGAGTCGGCGCAAGTGGCTGACACCCCGGCTGGTGAAGCTG	921
QY	726	CCCGGGAGCCCTTGACACGACGACGACGAGGAGGAGCGACGAGCGGGCTGTCAGAGCGGGGG	785
Dp	922	CCGGAGGACCCCTTGAACCCAGAACCCAGGAGGAGGAGCGAGCGGGCTGTCAGAGCGGGGG	981
QY	786	ACCCCTCTACGGGCCCCCGCTGCAGATGTCGTCTCTCCACGATGGAGCGCTTCGCGGGCCT	845
Dp	982	ACCCCTCTACGGGCCCCCGCTGCAGATGTCGTCTCTCCACGATGGAGCGCTTCGCGGGCCT	1041
QY	846	GCTGCCCCGTGCTGGGCGACGCCATTCATCCGACGACATCCCGCAGCGGATCTGTGGCCGCTG	905
Dp	1042	gctgccccgttgtggtgccagcccatctaccgcaagcatcccgcaaggagcatctgtgcccgtg	1101
QY	906	GGCGACACGGCCCTCTGCGGGACCCATCTGCGGGCGGCGCTGAAACGGACCATCTCCGGCC	965
Dp	1102	ggcgacaagctccctctcgtggacccaacctggtggcgagctcgtgaagccaacctccggcc	1161
QY	966	GGCGTTCCGGCGGGAAGTGGAGAAGACAGCGCTTCCCTTAGCGCAAGAAAGCCCGGAGAT	1025
Dp	1162	ggcgttccgggtccgggaagtgagaagagaagacgctgtctctaaaggcaagaagcccgagat	1221
QY	1026	AGACGAGAGCCTCATCTTCTTACAGAAAGTGGAGCTGGAAGCCCTGCGTGATGCGGCCCT	1085
Dp	1222	agaagagagcctcatctctctacaagaagtgtgagctgtgaagcctgtgtgaltgtgcgacct	1281
QY	1086	GCTGGCACCCAGATGGACGGCGGTGAAGCGCATCCCTTCACCTGACGACGACTGGAGCT	1145
Dp	1282	gctggcacccagatggacggcggtgaagcgcatcccttcacctaagagcagcttgaagct	1341
QY	1146	CCTAAACGATAACTGGATGAGCTTACCCACAGGTTACCCCGAGTGTGATCATCACCA	1205
Dp	1342	ctctaaagcataaactgtgaigtctctaccacaagaagtctaccocgagctgtgtgaacctaga	1401
QY	1206	CCTGGGCTACCTCTTCTCTCAAGATGAGCCCTGAGACATTCGCAAGTGGAATGTGAGCTC	1265
Dp	1402	ctgtggtactacctctctctcaagaatggacccctggagacatctgcaagtgtgaatgtgaagctc	1461
QY	1266	CCTGAGACCCCTGAAAGCTTGTGTTGAAGTCGACAAAGGGCAACGAATGATGTCACAGC	1321
Dp	1462	ccgtgagacccctgaaagcttgtctgtcgtgaagtcacaacaaaggacagaaatgtgacctc----	1516
QY	1326	TCTCTGGCGGCGCCCTCCACAGGTGGCCACCTCGATGACGCCCTTGTGTAAGGAAGGGG	1385
Dp	1517	-----caggtgtggcaacctgtatcgacgaccttgtgtgaaggaaggaagg	1557
QY	1386	CGACGTAGACAAAGACACCCCTAGACACCCCTGACCGCCTTCTACCTGGGTACTGTGCTC	1445
Dp	1558	cgagctagacaagaacacaccttagacacctgtacgacctctctacccttgggttacctgtgtct	1617
QY	1446	CCTTAGGCCCCGAGAGAGCTGAGCTCCGTGCCCCCCACAGACATCTGGGCGGTAGAGCCCA	1505
Dp	1618	ccctagagcccgagagatgtgagctcgtgtgcccccaagcatctgtggcgtgtcaaggccca	1677
QY	1506	GGACCTTGACACGCTGATACCCCAAGGACGAGCTGGAGCTGCTCTATTCCCAAGGCCCGCTTGC	1565
Dp	1678	ggacctgtgaacagtgtgtaacccaaggacagctgtgaagcttccctatcccaaggcccgctgtgc	1733
QY	1566	TTTTCCAGAACATGAACGGGTCCGAATATCTTGTGAAGATCAAGTCCCTTCTGCGGTGGGC	1625
Dp	1738	tttccagaacatlgaaacggtgcgcgaatactgtggaagatccagctctctctgtgtgtgggc	1797
QY	1626	CCCCACGGAGGATTTGAAGCGCTCACTAGCAGCAATGTGAGCATGTGACTTGGCCACGTT	1688
Dp	1798	ccccacggagagatttgaagcgtctcagctcagcaaatgtgaacatltgactltgtgccacgt	1857
QY	1686	CATCAAGCTGGGAGCGGATCGCGGCTGCCGCTTGATGCTGGGCTGAGAGGTCAACAACTTCT	1745
Dp	1858	catgaagcttggagacggaatgtggtgtgtctgcgcgttgttactgtgtgcgtgaagtgtgcagaactctc	1917

[illegible]

```

1  RESULT: 9
2  US-09-698-010-11861
3  : Sequence 11861, Application US/09698010
4  : GENERAL INFORMATION:
5  :
6  : APPLICANT: Williamson, Mark
7  :
8  : TITLE OF INVENTION: NOVEL NUCLEIC ACID MO
9  :
10 : FILE OF INVENTION: THEREFOR
11 :
12 : FILE REFERENCE: 1600.2029-001
13 :
14 : CURRENT APPLICATION NUMBER: US/09/698.010
15 :
16 : PRIOR FILING DATE: 2000-10-27
17 :
18 : PRIOR APPLICATION NUMBER: 60/162.358
19 :
20 : PRIOR FILING DATE: 1999-10-29
21 :
22 : NUMBER OF SEQ ID NOS: 15684
23 :
24 : SOFTWARE: FastSeq for Windows Version 4.0
25 :
26 : SEQ ID NO 11861
27 :
28 : LENGTH: 2289
29 :
30 : TYPE: DNA
31 :
32 : ORGANISM: Homo sapiens
33 :
34 : US-09-698-010-11861

```

Query Match	92.5%	Score 1976.8	DB 27	Length 2289
Best Local Similarity	97.3%	Pred. No. 0		
Matches 2053	Conservative	0	Mismatches 27	Indels 30
				Gaps 3
QY	12 TGGCCGGCCACATCCCGTGTGTGTATCAGCGCGCGAGAGAGAGAGCTTACCGGTGAGCCACGGT	71		
Db	202 TGGCGGCGCACCCCGCTGTCTGT	261		
QY	72 GCGTCCCTCCCTGGGATCTACACAGACCATGGCCCTTG-CACGGCTGCACCCCTGTTGGT	130		
Db	262 GCGTCCCTCCCTGGGATCTACACAGACCATGGCCCTTG-CACGGCTGCACCCCTGTTGGT	321		
QY	131 CCTGGGGGAGCGGCC-----TGGCAGGCTCTGTTTCCGTCGCTTCAGGCTGGAGTGGGT	185		
Db	322 GTCTGT	381		
QY	186 GCATCCCGCGAGAGACCTGGCTGT	245		
Db	382 GT	441		
QY	246 CCTGCACACCCCGCATACATTTCCACGCTTCCCTCGCCCACTCTTGGCTTCCCGT	305		









```
OY 2106 CCTGCAGACA 2115
|||||
Db 2278 cctgcagaca 2287

RESULT 11
US-09-710-280-2725
: Sequence 2725, Application us/09710280
: GENERAL INFORMATION:
: APPLICANT: Holtzman, Douglas A.
: APPLICANT: Geating, David P.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: FILE REFERENCE: 1600.2007-001
: CURRENT APPLICATION NUMBER: us/09/710,280
: PRIOR FILING DATE: 2000-11-10
: PRIOR APPLICATION NUMBER: 60/164,537
: NUMBER OF SEQ ID NOS: 2763
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2725
: LENGTH: 2289
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-710-280-2725

Query Match 92.5%; Score 1976.8; DB 28; Length 2289;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 2053; Conservative 0; Mismatches 27; Indels 30; Gaps 3;

OY 12 TGGCGGCGCATCCCGTGTGCTGAGCGGAGACAGAGACTACCGGTGAGCCAGCGT 71
|||||
Db 202 tggcgcgacatcccggtgtgctgtagcgcgagacagagactacggtgagccagcgt 261

OY 72 GCTTCCTCCTCGGATGTACACAGACCATGCGCTTG-CAACGGCTCGACCCCTTTGGT 130
|||||
Db 262 gcttcctcctcggtgtgtacacagaccatgcttgcaacggtcgagccctgtgg 321

OY 131 CCTGGGGGACCGCC-----TGGCAGCGTCTGTCTCTGCTTTCACCTCGATGGGT 185
|||||
Db 322 gtccgtggagcccgccctcgagacgctctgtctctcagcctcgagtgggt 381

OY 186 GCATCCCGGAGAGACCTCGGCTGAGAGACAGAGACGAGTCTGCCCTGGGGGAGT 245
|||||
Db 382 gcagccctcgagagaccgctgtgagagacagagagagagctgcccctggagaggt 441

OY 246 CCTGACAACCCCCATAATTCAGCTCTCCCTCGCAACTCTTGGCTTCCGCTG 305
|||||
Db 442 cctgccaacccacataatcattccagctctccctcgcaactccttggcttccgtg 501

OY 306 TGGGAGGTGTCGGGCTGAGCAGGAGCGTCCGGAGCTGGGTGAGCTTGGACACA 365
|||||
Db 502 tgggaggtgtcgggctgagcagagcgtgtccggagctgggtgagcttggacaca 561

OY 366 GAAGAATGTCAAGTCTCAAGAGACAGACTGCGTGTGCTGCTACCGGCTCTGAGCC 425
|||||
Db 562 gaagaatgtcaagtctcaagagacagactgctgtgtgtgtcaccggctctctgagcc 621

OY 426 CCGCGAGAGACTGAGCGCCCTCCATTGGACCTGCTGCTATTTCCTCAACCCAGATGCTT 485
|||||
Db 622 cccgagagactgtgagcgccctccattggacctgtgtatctcccaaccagatgagct 681

OY 486 CTCGGGGGCGGAGCGCTGACCCGTTTCTTCCCGCATCACGAAGCGCAATGTGAGCT 545
|||||
Db 682 ctccggggcgagcgctgacccgcttctctccgcatacgaagcgaatgtgacct 741

OY 546 GCTCCCGGAGGGGCTCCCGAGCGACAGGCGTGTGCTGCTGCGGCTGCGCTGCTGGGG 605
|||||
Db 742 gctcccgaggggctcccgagcgacagcggctgtgctgtcggtcgtgacctgtgggg 801

OY 606 TGTGCGGGGTCTCTGTGAGCGAGGCTGATGTGCGGGCTCTGGGAGCGCTGGCTTGCGA 665
|||||
```

```
Db 802 tgtgggggtgtctgtcgagcgagcgtgatlgtcggtgtcttggaggccgtgttgcga 861
OY 666 CCTGCTGGGCGGCTTTGTGGCGGAGTCCGGCGAAGTGTGTACCCCGGCTGTGAGCTG 725
|||||
Db 862 ctgtccttggcgcttltgtggcgagctcgccgaagtgtgtctaccccggtcgtgtgagctg 921

OY 726 CCGGAGACCCCTGAGACGAGACAGCAGAGGAGCGAGCGGCTGTGAGGCGGGGG 785
|||||
Db 922 ccgggagcccttgaacacagagcccgagagcgagcgagcgagcgagcgagcgagcgagcg 981

OY 786 ACCCCCTACGAGCCCCCGGTGACATGCTGTCTCCACGATGGAAGCTGTGCGGGGCT 845
|||||
Db 982 accccctacgcccccgctgacatgtgtcttccacagatgaagcgtctcggtggcct 1041

OY 846 GCTGCGCGTGTGGGCGGCGGCGGATCATCCGAGAGATCCCGAGGCGATCGGCGGCTG 905
|||||
Db 1042 gtgtccggtgtgtggcgagccatcatccgagagatcccgagagcattgtggcggtg 1101

OY 906 GCGCAACGCTCTCTCTCGGAGCCCATCTGCGCGAGCGCTGAGAGCATCTCTCGGCG 965
|||||
Db 1102 ggggcaacgctctctctgtggagccatctctgtggcgagcgtgaagagacatccctcggtc 1161

OY 966 GCGTTCGGGCGGAGAGTGAAGACAGAGCTGTCTCTTACAGGCAAGAGCGCCGAGAT 1025
|||||
Db 1162 ggggttcggcgggaggtggagagacagcgtgtcttcaaggcaagagcgccgagat 1221

OY 1026 AGACGAGAGCTCATCTTCTCAAGAAATGGAGACTGGAACCTGCGTGTGAGCGCCCT 1085
|||||
Db 1222 agacgagagctcatctctcaagaagtggagctgtgaagcgtgtgtgtgtgtgtgtgtgt 1281

OY 1086 GCTGGCCACCCAGATGAGCCCGGTGAACGAGCATCCCTTCACTACGAGCGAGTGAAGT 1145
|||||
Db 1282 gtgtgcccacagatgtgacggtgtgaagcgtatccctcaactcaagagcgtgtgagct 1341

OY 1146 CTTAAAGCATTAAGTGTGATGAGCTTACCCACAAAGTTTACCCCGAGTGTGATTCAGCA 1205
|||||
Db 1342 cctaagaataaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1401

OY 1206 COTGGGCTACCTCTTCTCAAGATGAGCGGCTGAGAGCATTTGGCAAGTGGATGAGCTC 1265
|||||
Db 1402 cctgggtcaactctctcccaagaatgaagcgtgtgaagcgtgtgtgtgtgtgtgtgtgt 1461

OY 1266 COTGAGACCTGAAAGGCTTTGCTTGAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGG 1325
|||||
Db 1462 cctgagagaccttgaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1516

OY 1326 TCTTCGGCGGCGCTCCACAGGTGCGACCTGATCGACCGCTTTTGAAGGAGGAGG 1385
|||||
Db 1517 -----caggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1557

OY 1386 CCAGTTAGCAAAAGACACCTAGACACCTGACCGCTTACCTGAGGTACCTGTGCTC 1445
|||||
Db 1558 ccaagtaaacaaagacacacacagagacccgtgacgctctcaactctgtgtgtgtgtgt 1617

OY 1446 CCTCAGCCCCGAGAGGCTGAGCTCGTGTGCCCCCGACAGCATCTGGGGGTCAGGCCCCA 1505
|||||
Db 1618 cctcagccccgagagagctgtgaccccaagcagctgtgtgtgtgtgtgtgtgtgtgtgtgt 1677

OY 1506 GGACCTGAGACGTTGTGACCCCAAGGACGCTGAGCTCTATCTCAAGGCGCCGCTTGG 1565
|||||
Db 1678 ggaactgtgacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1737

OY 1566 TTTTCAGAACTGAAGCGGCTCGAATACTTGTGTAAGATCCAGTCTTCTCGGTGGGGC 1625
|||||
Db 1738 ttccagaacatagaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1797

OY 1626 CCCCAGGAGATTGTAAGGCGCTCAGTACAGAGATGTGAGCATGTGAGCTTGGCGACATT 1685
|||||
Db 1798 ccccaagagagattgtgaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1857

OY 1686 CATGAAGCTGCGAGCGATGCGGTGCTGCTGCTTGAATGTGTGAGTGAAGTGAAGTCTTCT 1745
|||||
```



```
QY 1326 TCCTGCGGCGCCCTCCACAGATGCGCCACCTGATCGACGCTTTGTGAAGAGAGGGG 1385
Db 1517 -----cesgtgycacacccctgatcgaccgcttctggaagggaagg 1557
QY 1386 CCAGCTAGACAAGACACCTTAGACACCTTAGACCGCTTCTACCTTGGGTACTGTGCTC 1445
Db 1558 ccaagtagacaagaacacccctagacacccctgacgcgctcttacccttggtaacctgtctc 1617
QY 1446 CCTGACGCGCCGAGAGATGAGCTCCGTGCCCCCAAGACGATCTGCGGGTACAGGCCCA 1505
Db 1618 cccagccccagagagctgagctccgtgcccccaagcagcaltctggcgctcaaggccccca 1677
QY 1506 GGACCTGAGACAGTGTGACCCCAAGGACAGCTGAGCTCCTCTATCCCAAGGCCCGCTTGC 1565
Db 1678 ggaacctgacacgctgtgacccaaggcagcttgaaagctcttaccacaaagcccgcttgc 1737
QY 1566 TTTTCAGAACATGAACGGGCTCCGAATCTTCTGTGAAGATCCAGTCTTCTGAGTGGGGC 1625
Db 1738 ttccagaaacatgaacggtctcgaatactctgtgaagatccagctcttcccttgggtgggc 1797
QY 1626 CCCACGAGAGATTTGAAGGGCGCTCAGTCAAGACAATGTGACATGAGCTTGGCCACTT 1685
Db 1798 cccacaggaagatctgaaggcgctcagcagaatgtgaagcatggaacttgccagc 1857
QY 1686 CATGAAGCTCGGACGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1745
Db 1858 catgaagctcggaagcgagctggtgctgctgctgctgctgctgctgctgctgctgctgct 1917
QY 1746 GGGACCCACCTGAGAGGCTTGAAGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1805
Db 1918 gggacccacactgagagggcttgaaggcgagagagcgagagcgagcgagcgagcgagctg 1977
QY 1806 CCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1865
Db 1978 cctagggagggagggagggagggagggagggagggagggagggagggagggagggagggag 2037
QY 1866 CAACGCTACTGTGCTCAGACCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCT 1925
Db 2038 caacgctactgtgtctcagacctcagctcagctcagctcagctcagctcagctcagctcag 2097
QY 1926 CCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1985
Db 2098 cctagggagggagggagggagggagggagggagggagggagggagggagggagggagggag 2157
QY 1986 AGGCGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2045
Db 2158 aggcgcacaccccttctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 2217
QY 2046 GCACGGGTGATCCCGTTCACCCCAAGAGAACTCGCGCTCAGTAAACGGAGACATGCC 2105
Db 2218 gcaacgggtgatcccggttcaccccaagagaactcgctcagtaaacggagaacatgccc 2277
QY 2106 CCTGCAGACA 2115
Db 2278 cctgcagaca 2287
```

## RESULT 13

```
US-09-710-286-2845
; Sequence 2845, Application US/09710286
; GENERAL INFORMATION:
; APPLICANT: Geating, David P.
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2005-001
; CURRENT APPLICATION NUMBER: US/09/710.286
; PRIOR APPLICATION NUMBER: 60/164.255
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4115
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 2845
; LENGTH: 2289
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-710-286-2845
```

Query Match 92.5%: Score 1976.8; DB 28; Length 2289;  
Best Local Similarity 97.3%; Pred. No. 0;  
Matches 2053; Conservative 0; Mismatches 27; Indels 30; Gaps 3;

```
QY 12 TGGCGGCGACCTCCCGTCTGCTGATGACGCGGAGACAGAGACTACCGGTGAGCCACAGT 71
Db 202 tggcgcgacatcccgctctgtgtgacgcyggaagagagctacacgttggaccacagct 261
QY 72 GCTTCCTTCCTGGAGATACACAGACCATGAGCCTTG-CAACGGCTGACCCCTGTTGCT 130
Db 262 gcttccttcctggagatcacagaccatgtgcttgcaaacggtctgagccctgttgg 321
QY 131 CCTGCGGAGCGGCC-----TGGCAGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 185
Db 322 gtccgtggagaccccgccctcgagacccctctgtctctctctcagcctcagatggct 381
QY 186 GCATCCCGGAGACCCCTGCGCTGAGAGACAGGACGAGTCTGCCCTGCGGGGAGT 245
Db 382 gcaagccctcgagacccctgctgagagagacagggcagagagctcgccccctggagagct 441
QY 246 CCTGACAAACCCCTTACCATTTTCCAGCTCTCCCTGCTGCCAATCTCTTGGCTTCCGCTG 305
Db 442 cctggacaaaccccttaccatttccagctctccctcgcacactcttgcctccgtg 501
QY 306 TGGGAGGAGTTCGCGGCTGAGACAGGAGCTGTCCGGAGGCTGAGCTGAGCTTGGACACA 365
Db 502 tgggagagtgctcgagcctcgagacagagagctgtccggagagctgtgtgcttggcaca 561
QY 366 GAAGATGTCAAGCTCTCAGACAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAGCC 425
Db 562 gaagatgtcaagctctcagacagagcagctgtgtctgtgtctcagcgtctctcagac 621
QY 426 CCGCGAGAGCTGAGAGCGGCTCCCATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 485
Db 622 cccgagagacctgagacgctcccatctgagctgtgtatctccacccagatgctgt 681
QY 486 CTGCGGCGCCAGAGGCTCAGCCGCTTCTCTCCCGCATCAGAAAGGCAATGTGAGACT 545
Db 682 ctgcyggccccagagccttgacacgcttctctcccgatcacagagcgaatgtggacct 741
QY 546 GCTCCCGAGGGGGGCTCCGAGCGACAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 605
Db 742 gctcccgagggggtctcccgagcgacagcgctgtgtctgtcgtgtgtgtgtgtgtgtgt 801
QY 606 TGTGCGGGGCTCTGCTGAGCGAGGCTGATGCTGGGCTCTGGGAGGCGCTGCTGCTGCTGCT 665
Db 802 tgtcgggggtctctgtcgtgagcgagctgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 861
QY 666 CCTGCTGAGGCGCTTGTGCGCGAGTCCGCGAAGTGTCTTACCCCGGCTGCTGAGAGCTG 725
Db 862 cctgctgagcgcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 921
QY 726 CCGGAGACCTTGGACAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 785
Db 922 ccggagaccccttgagacagagcagcgagcgagcgagcgagcgagcgagcgagcgagcg 981
QY 786 ACCCCCTAGGGCCCCCGGTGACATGCTGCTCCACAGATGAGCGCTGCGGGGCGCT 845
Db 982 accccctagggcccccggttgacatgtctctccacagatgagcgtctcgaggggcgct 1041
QY 846 GCTGCGGCTGCTGAGGAGCCCATCATCCGAGCATCCCGAGGAGCATCTGAGCGAGCTG 905
Db 1042 gctgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1101
QY 906 GCGGCAACGCTCTCTGCGGAGCCATCTGCGGAGGCTGAGGAGCATCTCTCGGCG 965
```

[illegible]

OY	2046	GCACGGGATGCCCGTTCACCCAGACAAGACTGGCTCAGTAAGAAGGAAATGCCC	2105
Db	2218	gcacggggtggtcccgcttccaccacaagaactcgcgtcagtaaacggaacatgccc	2277
OY	2106	CCTGCAGACA	2115
Db	2278	cctgcagaca	2287
RESULT 14			
US-09-726-171-1716			
: Sequence 1716, Application US/09726171			
: GENERAL INFORMATION:			
: APPLICANT: Geating, David P.			
: APPLICANT: Holtzman, Douglas A.			
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES			
: TITLE OF INVENTION: THEREFOR			
: FILE REFERENCE: 1600.2015-001			
: CURRENT APPLICATION NUMBER: US/09/726.171			
: CURRENT FILING DATE: 2000-11-29			
: PRIOR APPLICATION NUMBER: 60/167,845			
: PRIOR FILING DATE: 1999-11-29			
: NUMBER OF SEQ ID NOS: 2579			
: SOFTWARE: FastSeq for Windows Version 4.0			
: SEQ ID NO 1716			
: LENGTH: 2289			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
US-09-726-171-1716			
Query Match 92.5%; Score 1976.8; DB 29; Length 2289;			
Best Local Similarity 97.3%; Pred. No. 0;			
Matches 2053; Conservative 0; Mismatches 27; Indels 30; Gaps			
OY	12	TGGCCGGCACTCCCGTGTGTGTGACCGCGCGAGAGAGCTACCGTGACCACGCT	71
Db	202	tggccggcaactcccgctgtgtgtgacgcygcagagagtctacggttgagccacggt	261
OY	72	GCCATCCCTCCGGGATTTACAGACCATGGCCATTG--CAACGGCTGCACCCCTTTGGT	130
Db	262	gccatccctccgggatattacagaccatggccttgcacaaagtcgcgacccctgtggg	321
OY	131	CCTGTGGGACCGCCCC-----TGCGAGCCTCTCTTCTCGCTTCACACTTCGATGGGT	185
Db	322	gtcctgttggagaccgccgcctcgcgagacctctgttctccttcagctcgtgagttggt	381
OY	186	GCATCCCGCGAGAGCCTGTGCTGTGAGAGACAGGACGAGAGTGTGCCCCCTTGCGGGAGT	245
Db	382	gcagcccttcgagagaccctgcgtgagagagacagtgagagtgctgcgcccttgcgaggt	441
OY	246	CCTGCACACCCCCCAATACATTTCAGGCTCTCCGCTTGCCCAACTCTTGGCTTCCGCG	305
Db	442	cctlgccaaccaactaacatttccagccttccccctgcgcaactcctltgcctcgtg	501
OY	306	TGCGAGGTGTCCGCTCTGAGCACGAGAGCTGTCCGCGAGCTGAGCTGTGACCTTGACCA	365
Db	502	tgcgaggtgttgcgcgcctgcgagacgagcgtgtgtcgggagagctggtgttgccttgcaca	561
OY	366	GAGAATGTGAAGTCTCTAACACAGAGCAGCTGCGCTGTCTGACTCACCGGCTTCTGAGCC	425
Db	562	gaaagtattcaagctctcaacagagcagctgcgtgtctgtgtctcacaggtcttctcgaagc	621
OY	426	CCCGAGAGACCTGAGAGGCGCTCCATTCGAGCTGCTGGATTCCTCAACCCAGATGCGTT	485
Db	622	cccgaggagacttgagagccctcccatitgagcttgtctatatccccaacccaagaggtt	681
OY	486	CTCGGGGCCCAAGGCTGCACCCGTTTTCTTCTCTCCGATCAGAAAGCCAAATGTGAGCT	545
Db	682	ctcgggccccagcgtgcacccgttcttctccgcgtatcagaaagccaatgtgacct	741
OY	546	GCTCCCGAAGGGGGCTCCCGAGCAGACAGCGCTGTCTGCTGTGGGGCTCTGCTGGGG	605



Db 241 gagccctgagacatcgcgaagtgaaatgtaacgtccctgagacccctgaaagcttgc 300  
QY 1290 TGAAGTGACAAAGGAGCAGAAATGAGTCTCTCAGGCTCTCGCGGCGCCCTCCACAGGT 1349  
Db 301 tgaagtaacaaaaggacgaaatgagtcct-----caggt 336  
QY 1350 GGCCACCCTGATGCACGCGTTTGTGAAGGGAAGGGCCAGCTAGACAAGACACCTAGA 1409  
Db 337 ggcacacctatgcagccgttctgaaaggaaaggccagctagacaaagacacctaga 396  
QY 1410 CACCTGACCGCCTT-CTACCTGGGTACTGTGTCTCTCAGGCCCGCAGAGAGCTGAGCT 1468  
Db 397 caccctgaccgctctgtacccttgtaacctgtctcctcagcccgagagctgagct 456  
QY 1469 CCGTGCCCCCAGAGAGATGTGGCGGTACAGGCCCAAGAGACTGGACAGCTGTACCCAA 1528  
Db 457 ccgtgccccccagacatctggtggtggtggtggtggtggtggtggtggtggtggtggt 516  
QY 1529 GGCAGCTGAGAGTCTCTATCCCAAGGCCCGCCTTGTCTTCCAGACATGAAACGGTCCG 1588  
Db 517 ggcagctgagacgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 576  
QY 1589 AATACTTGTGAAGATCCAGTCTTCTCTGGGTGGGCCCCCAGCAGAGATTGAAGCGC 1648  
Db 577 aatctctgtgaagatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 636  
QY 1649 TCAGTCAAGCAATGTAGCATGTGACTTGGCCACGTTTATGAACTGGGAGGATGCGG 1708  
Db 637 tcagtcagcagaatgtgagcatggaacttggtcctcctcctcctcctcctcctcctcctc 696  
QY 1709 TGGCGCGTGTGACTGTGAGTGTGAGTGTGAGAACTTGTGGAGCCCGCAGTGAGGGCGCTGA 1768  
Db 697 tgcgcgctgtgacgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 756  
QY 1769 AGCGGAGGAGCGGACCGCGCGCGGTGCGGACTGATCTTACGGCAGCGCAGGAGCACC 1828  
Db 757 agcgagagagcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 816  
QY 1829 TGGACAGGCTGGGGCTGGGGCTACAGGGCGGCAATCCCAAGGCTTACTGCTTACAGC 1888  
Db 817 tggacagcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 876  
QY 1889 TCAGCTGCAAGAGACCTCTCGGGGACGCCCTGCTCTAGACCTGACCTGTCTCA 1948  
Db 877 tcagcagcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 936  
QY 1949 CCGTCTGGGCACTGCTCTTACCTTCAACCTG-CCCTGAGGGCCCACTCCCTTGTGGC 2007  
Db 937 ccgtcctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 996  
QY 2008 CCCAGCCCTGTGGGGATCCCGGCTGGGCGCAGAGGACAGGAGGATGATCCCGTTCCAC 2067  
Db 997 gtcagccctgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1056  
QY 2068 CCCAAGAGAACTCGGCTCAGTAAGGGAACATGCCCGCTGCAGACA 2115  
Db 1057 cccaagaaactcgtcctcagtaaacggaacatgccccctgcagaca 1104

Search completed: December 7, 2001, 10:40:45  
Job time: 2818 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2001, 09:48:27 : Search time 350.8 Seconds  
(without alignments)  
8403.290 Million cell updates/sec

Title: US-09-215-035-1  
Perfect score: 2138  
Sequence: 1 AGCAATTCGCGTGGCGCGCC.....AAAAAAAAAAAAAAAAAAAA 2138

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 906424 seqs, 689399951 residues

Total number of hits satisfying chosen parameters: 1812848

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_NA\_New:\*  
1: /cgn2\_6/ptodata/1/pna/PCN\_NEW\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/pna/US06\_NEW\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/pna/US09\_NEW\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	425.2	19.9	562	US-09-828-792-48	Sequence 48, Appl
2	425	19.9	553	US-09-828-792-182	Sequence 182, App
3	357.8	16.7	442	US-09-828-792-151	Sequence 151, App
4	219.8	10.3	396	US-09-970-966-126	Sequence 126, App
5	52.8	2.5	2511	US-60-325-795-1	Sequence 1, Appl
6	52.2	2.4	4282	US-09-778-927A-15	Sequence 16, Appl
7	52	2.4	19363	US-09-815-264-67667	Sequence 67667, A
8	49.8	2.3	17466	US-09-815-264-58041	Sequence 58041, A
9	49	2.3	11720	US-09-760-446A-2456	Sequence 2456, App
10	48.6	2.2	17403	US-09-815-264-72273	Sequence 72273, A
11	47.8	2.2	50229	US-09-815-264-76667	Sequence 76667, A
12	47.8	2.2	76304	US-09-815-264-76673	Sequence 76673, A
13	47.6	2.2	49792	US-09-815-264-80624	Sequence 80624, A
14	46.8	2.2	1926	US-09-815-264-79254	Sequence 79254, A
15	46.4	2.2	1926	US-09-249-585A-2	Sequence 2, Appl
16	46	2.2	71989	US-09-724-876-2	Sequence 4342, App
17	46	2.2	1206	US-60-325-448-4342	Sequence 4342, App
18	45.8	2.1	471	US-60-325-448-4189	Sequence 4189, App
19	45.8	2.1	771	US-09-815-264-82984	Sequence 82984, A
20	45.8	2.1	44821	US-09-815-264-82984	Sequence 82984, A
21	45.6	2.1	773	US-09-815-264-68091	Sequence 68091, A
22	45.4	2.1	773	US-60-325-448-4341	Sequence 4341, App
23	45.4	2.1	2269	US-09-243-560B-1	Sequence 1, Appl
24	45.4	2.1	16641	US-09-815-264-66356	Sequence 66356, A
25	45.4	2.1	108592	US-09-815-264-82756	Sequence 82756, A
26	45.2	2.1	2824	US-09-815-264-82420	Sequence 82420, A
27	45.2	2.1	3155	US-09-822-846-23	Sequence 23, Appl

28	45.2	2.1	3155	US-09-822-846A-23	Sequence 23, Appl
29	45.2	2.1	3619	PCT-US01-27760-93	Sequence 93, Appl
30	45.2	2.1	12748	US-09-815-264-67258	Sequence 67258, A
31	45.2	2.1	45140	US-09-815-264-64392	Sequence 64392, A
32	45	2.1	19036	US-09-815-264-77573	Sequence 77573, A
33	44.6	2.1	108634	US-09-815-264-62252	Sequence 62252, A
34	44.4	2.1	374	US-09-388-906A-17125	Sequence 17125, A
35	44.4	2.1	418	US-09-388-906A-10829	Sequence 10829, A
36	44.4	2.1	480	US-09-388-906A-20855	Sequence 20855, A
37	44.2	2.1	6238	US-09-815-264-82752	Sequence 82752, A
38	44	2.1	26111	US-09-815-264-62575	Sequence 62575, A
39	44	2.1	38072	US-09-815-264-65084	Sequence 65084, A
40	43.8	2.0	4094	US-09-760-446A-523	Sequence 523, App
41	43.6	2.0	3374	US-09-815-264-73824	Sequence 73824, A
42	43.6	2.0	5076	US-09-815-264-82390	Sequence 82390, A
43	43.6	2.0	17092	US-09-815-264-82641	Sequence 82641, A
44	43.6	2.0	42999	US-09-799-462A-17	Sequence 17, Appl
45	43.4	2.0	479	US-09-388-906A-23028	Sequence 23028, A

## ALIGNMENTS

```
RESULT 1
US-09-828-792-48
: Sequence 48, Application US/09828792
: GENERAL INFORMATION:
: APPLICANT: Geating, David P.
: APPLICANT: Robison, Keith E.
: TITLE OF INVENTION: Nucleic Acid Molecules Derived from a
: FILE OF INVENTION: Human Fetal Lung Library
: FILE REFERENCE: MN98-34pa
: CURRENT APPLICATION NUMBER: US/09/828,792
: PRIOR FILING DATE: 2001-04-09
: PRIOR APPLICATION NUMBER: US/09/330,360
: PRIOR FILING DATE: 1999-06-11
: PRIOR APPLICATION NUMBER: 60/090,258
: PRIOR FILING DATE: 1998-06-22
: NUMBER OF SEQ ID NOS: 1102
: SOFTWARE: PastSeq for Windows Version 3.0
: SEQ ID NO 48
: LENGTH: 562
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (133)...(562)
: NAME/KEY: misc_feature
: LOCATION: (1)...(562)
: OTHER INFORMATION: n = A,T,C or G
US-09-828-792-48

Query Match 19.9%; Score 425.2; DB 5; Length 562;
Best Local Similarity 93.7%; Pred. No. 2e+82;
Matches 487; Conservative 0; Mismatches 25; Indels 8; Gaps 4;

OY 12 TGCGCGGCATCCGCTGCTGTGAC-GCGCGACAGAGACTACCGGTGACCCACG 70
DB 44 tgcgcgcacacccctgctgtgtacagcgcgacagagactaccggtgacccacg 103
OY 71 TGCCCTCCCTCCCTGGAGCTACAGACCAAGGCTTG-CAACGGCTGACCCCTGTGG 129
DB 104 tgcctccctcccttggagactacacagacacagccttccaaagcctgacccctgttgg 163
OY 130 TCCTGTGGGAGACCGCC-----TGCGACCTCTCTTCTTCTTCAAGCTCGATGGG 184
DB 164 gttctctgtgggaccccgccctcgcgcagcctctcttctcagcctcgagatgg 223
OY 185 TGCAATCCCGGAGAGACCTGTGCTGAGAGACAGAGAGAGAGTCTCCCTCGGGGAG 244
DB 224 tgcagccctcagagacccctgtgtgagagacagagcagagagctgacccctgagag 283
```



```
Db 249 attccagcctctccctccgcaacctctcgtcctccgctgctgagggagctgagcgc 308
|
QY 325 AGCAGGAGAGGTGTCCGGAGCTGGCTGTGGCTTGACAGAGAATGTCACTCTCA 384
|
Db 309 agcacggagctgtcccgagctgctgctgctgctgcaagaagaatgtcaactc 368
|
QY 385 ACAGAGAGCTGGCTGTGGCTCACGGCTCTGAGCCCGCCGAGGAGCTGAGCGCC 444
|
Db 369 acagagagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 428
|
QY 445 CTCCTATTTGAGC 457
|
Db 429 ctccattgagc 441
|
RESULT 4
US-09-970-966-126
; Sequence 126, Application US/09970966
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Mollesh, David Alan
; APPLICANT: Fling, Steven P.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C6
; CURRENT APPLICATION NUMBER: US/09/970,966
; CURRENT FILING DATE: 2001-10-02
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 126
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-970-966-126

Query Match 10.3%; Score 219.8; DB 5; Length 396;
Best Local Similarity 80.9%; Pred. No. 3,5e-38;
Matches 279; Conservative 0; Mismatches 42; Indels 24; Gaps 1;

QY 1243 ATTGCAAGTGAATGTGACGTCCCTGAGAGACCTGTAAGCTTTGTAAGTGCACAA 1302
|
Db 8 actcgcaagtgaatgtgacgtccctgagagacctgagactgtgctgagtaacaa 67
|
QY 1303 GGGCAGCAATGAGTCTCAGGCTCTCGGCGGCCCTCCACAGGTGGCCACCTGATC 1362
|
Db 68 gggcagcaaatgagtcct-----cagtgagcaacctgac 103
|
QY 1363 GACCGCTTTGTAAGGAGGAGGCGACGTAGACAACACACCTGAGACCTGACCGCC 1422
|
Db 104 gacgcctctgtaaggaagggagcagctagacaagaccctagacacctgacgc 163
|
QY 1423 TTTCACCTGGGTACCTGTCTCTCCCTCAGCCCGGAGAGAGTGAAGTCTGCCCCAGC 1482
|
Db 164 ttctaaccttgtaacctgtctcctcctcagcccgagagagctgagctcgcccccaagc 223
|
QY 1483 AGCATCTGGGCGGTGACGGCCGAGACCTGAGACAGTGTACCCAGGAGGAGCTGACGTC 1542
|
Db 224 agcatctggcggtgtaagcggccagacctgtaacagctgagctgtaacagggcg 283
|
QY 1543 CTGTATCCCAAGGCGCCCTTGCTTTCCAGAACATGAACGGGTCC 1587
|
Db 284 atcccaagggctacctgctcctagacctcagcatgcaagagcc 328
|
RESULT 5
US-60-325-795-1
; Sequence 1, Application US/60325795
; GENERAL INFORMATION:
; APPLICANT: Su, Eric W
; APPLICANT: Wang, He
```

```
; APPLICANT: Zhi, Yu
; TITLE OF INVENTION: NOVEL SECRETED PROTEINS AND THEIR USES
; FILE REFERENCE: P-14990
; CURRENT APPLICATION NUMBER: US/60/325,795
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2511
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18)..(2306)
; OTHER INFORMATION:
US-60-325-795-1

Query Match 2.5%; Score 52.8; DB 6; Length 2511;
Best Local Similarity 50.6%; Pred. No. 0.073;
Matches 157; Conservative 0; Mismatches 147; Indels 6; Gaps 1;

QY 504 CACCGTTTCTTCTCCGCATCAGAGGCGCAATGTGACCTGCTCCGAGGGGGCTCC 563
|
Db 860 caccctgacctccgcyggaaccagctagacacctgccccgctgagggccggcga 919
|
QY 564 CGAGCGACAGCGCTGCTGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCTGCT 623
|
Db 920 gctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 979
|
QY 624 GAGCGAGCTGATGTGCGGCGCTGAGGAGGCGCTG-----GTTGCAAGCTGCGCTGCGCG 677
|
Db 980 actgagatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1039
|
QY 678 CTTTGTGCGGAGTGGCGCGAAGTGTCTACCCCGCTGCTGAGCTGCGCGGAGACCCCT 737
|
Db 1040 cctgcggcgagctgctgagcgccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1099
|
QY 738 GAGCAGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 797
|
Db 1100 ggcgaggaaggaaggaagctggaagagcgcgctgctgagcgcccgcccgccctccgcg 1159
|
QY 798 CCCCCCGTGC 807
|
Db 1160 cggccctcgcg 1169
|
RESULT 6
US-09-778-927A-16
; Sequence 16, Application US/09778927A
; GENERAL INFORMATION:
; APPLICANT: KHOSRAVI, Ramt et al.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID AND AMINO ACID SEQUENCES AND NOVEL
; FILE REFERENCE: 2786-0160P
; CURRENT APPLICATION NUMBER: US/09/778,927A
; CURRENT FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: ID 134453
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: IL135341
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 4282
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(4282)
; OTHER INFORMATION: n = a,c,g,t any unknown or other
US-09-778-927A-16
```



Sequence 2456, Application US/09760446A  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PTZ68  
CURRENT APPLICATION NUMBER: US/09/760,446A  
PRIOR FILING DATE: 2000-01-16  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: 60/214,886  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 60/217,487  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,758  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,963  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/217,496  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,447  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/218,290  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/225,757  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/226,868  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/216,647  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,267  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/216,880  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,270  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/251,869  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/235,834  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/234,274  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/234,223  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/228,924  
PRIOR FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/224,518  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,369  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/224,519  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,964  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/241,809  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/249,299  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/236,327  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/241,785  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/244,617  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: 60/225,268  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,368  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/251,856  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/251,868  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/229,344  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/234,997  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: 60/229,343  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,345  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,287  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,513  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/231,413  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/229,509  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/236,367  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/237,039  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,038  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/236,370  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/236,802  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,037  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,040  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/240,960  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/239,935  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/239,937  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/241,787  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,474  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/246,532  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/249,216  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,210  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/226,681  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,759  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/225,213  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/227,182  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,214  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/235,836  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/230,438  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/215,135  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: 60/225,266  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/249,218  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,208  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,213  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,212  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,207

PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,245  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,244  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,217  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,211  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,215  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,264  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,214  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,297  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/232,400  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/231,242  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,081  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,080  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,414  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,244  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/233,064  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/233,063  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,397  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,399  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,401  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/241,808  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,826  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,786  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,221  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,475  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/231,243  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/233,065  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,398

Query Match 2.3%, Score 49; DB 5; Length 11720;  
Best Local Similarity 54.8%; Pred. No. 0.84;  
Matches 97; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 598 TCGTGGGCTGCGGGGCTCTCTGCTGACGAGCTGATGTCGGGCTCTCTGGAGGCTTG 657  
DB 9537 TGTAGGAGT 9596  
QY 658 GCTTCGACGCTGCGCTGCGGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 717  
DB 9537 GTGCGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9656  
QY 718 GTGAGCTCCCGGAGCCCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 774  
DB 9657 GGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9713

RESULT 10

US-09-815-264-72273/c  
Sequence 72273, Application US/09815264  
GENERAL INFORMATION:  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Cao, Yongwei  
APPLICANT: Dotson, Stanton B.  
APPLICANT: Koshi, Jeffrey M.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: McIninch, James  
APPLICANT: Wu, Wei  
TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof  
FILE REFERENCE: 38-21(51237)G  
CURRENT APPLICATION NUMBER: US/09/815,264  
CURRENT FILING DATE: 2001-03-23  
PRIOR APPLICATION NUMBER: US 09/620,392  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: US 09/702,134  
PRIOR FILING DATE: 2000-10-31  
NUMBER OF SEQ ID NOS: 109669  
SEQ ID NO 72273  
LENGTH: 17403  
TYPE: DNA  
ORGANISM: Oryza sativa  
US-09-815-264-72273

Query Match 2.3%, Score 48.6; DB 5; Length 17403;  
Best Local Similarity 45.4%; Pred. No. 1.2;  
Matches 174; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 587 CGGCTCTGCGCTGCTGCGGCTGTGCGGGGCTCTCTGACGAGGCTGATGTCGGGCTC 646  
DB 9101 CGCTAGGGGCGAGGACACTCGCGGGGCGCGCGGAGCTGTAGAGAGCGGGGGGCTTG 9042  
QY 647 TGGAGGCTTGGCTTGGCACTGCTGCGGCTTTGTCGCGAGTGGCCGAGTGTCTGC 706  
DB 9041 CGAAGTGGCGCGCGGCTGTGTAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8982  
QY 707 TACCGCGCTGT 766  
DB 8981 TCACGCTGT 8922  
QY 767 CGGCTCTGAGGCGGGGAGCCCTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 826  
DB 8921 AGCGAGCGCGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8862  
QY 827 TGGAGGCTTGTGGGGGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 886  
DB 8861 CGGATGAGCGGCGGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8802  
QY 887 AGGCAATGTGCGCGCGGCGGCAACCTCTCTGCGGAGCCATCTGCGGCGAGCCTG 946  
DB 8801 GTGAGCGCGCGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8742  
QY 947 AACGAGCATCTCTCGGCGCGG 969  
DB 8741 CCGGAGCGAGGAGGAGGCGGCGG 8719

RESULT 11  
US-09-815-264-76667  
Sequence 76667, Application US/09815264  
GENERAL INFORMATION:  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Cao, Yongwei  
APPLICANT: Dotson, Stanton B.  
APPLICANT: Koshi, Jeffrey M.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: McIninch, James  
APPLICANT: Wu, Wei  
TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof

```
FILE REFERENCE: 38-21(51237)G
CURRENT APPLICATION NUMBER: US/09/815,264
CURRENT FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: US 09/620,392
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: US 09/702,134
PRIOR FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 109669
SEQ ID NO 76667
LENGTH: 50229
TYPE: DNA
ORGANISM: Oryza sativa
US-09-815-264-76673
```

Query Match 2.2%; Score 47.8; DB 5; Length 50229;

Best Local Similarity 50.7%; Pred. No. 2.6; Mismatches 112; Indels 0; Gaps 0;

Matches 115; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
QY 1756 GTGAGAGGCTGAAGCGGAGAGGAGCGGCGCCGCTGCGGAGATCTACGCGAG 1815
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||
Db 7150 gtgttgagagagagagagagagagagagagagagagagagagagagagagag 7209
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1816 CGGAGAGACACCTGAGACAGCTGGGCTGGGCTACAGAGGAGGAGGAGGAGGCTAC 1875
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 7210 gaggcgagagagagagagagagagagagagagagagagagagagagagagag 7269
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1876 CTGTCTTACCTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1935
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 7270 tacttcaagtaaccccaagagagagagagagagagagagagagagagagagag 7329
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1936 GGAAGCTGCTTCAAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1982
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 7330 ggggagagagagagagagagagagagagagagagagagagagagagagagag 7376
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

## RESULT 12

US-09-815-264-76673

Sequence 76673, Application US/09815264

GENERAL INFORMATION:

APPLICANT: Boukharov, Andrey A.

APPLICANT: Cao, Yongwei

APPLICANT: Dotson, Stanton B.

APPLICANT: Koshi, Jeffrey M.

APPLICANT: Kovalic, David K.

APPLICANT: Liu, Jingdong

APPLICANT: McIninch, James

APPLICANT: Wu, Wei

TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof

FILE REFERENCE: 38-21(51237)G

CURRENT APPLICATION NUMBER: US/09/815,264

CURRENT FILING DATE: 2001-03-23

PRIOR APPLICATION NUMBER: US 09/620,392

PRIOR FILING DATE: 2000-07-19

PRIOR APPLICATION NUMBER: US 09/702,134

PRIOR FILING DATE: 2000-10-31

NUMBER OF SEQ ID NOS: 109669

SEQ ID NO 76673

LENGTH: 76304

TYPE: DNA

ORGANISM: Oryza sativa

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(76304)

OTHER INFORMATION: unsure at all n locations

US-09-815-264-76673

Query Match 2.2%; Score 47.8; DB 5; Length 76304;

Best Local Similarity 50.7%; Pred. No. 3;

Matches 115; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 1756 GTGAGAGGCTGAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1815

```
Db 51291 gtgttgagagagagagagagagagagagagagagagagagagagagagagag 51350
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1816 CGGAGAGACACCTGAGACAGCTGGGCTGGGCTACAGAGGAGGAGGAGGAGCTAC 1875
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 51351 gaggcgagagagagagagagagagagagagagagagagagagagagagagag 51410
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1876 CTGTCTTACCTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1935
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 51411 tacttcaagtaaccccaagagagagagagagagagagagagagagagagagag 51470
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1936 GGAAGCTGCTTCAAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1982
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 51471 ggggagagagagagagagagagagagagagagagagagagagagagagagag 51517
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

## RESULT 13

US-09-815-264-80624

Sequence 80624, Application US/09815264

GENERAL INFORMATION:

APPLICANT: Boukharov, Andrey A.

APPLICANT: Cao, Yongwei

APPLICANT: Dotson, Stanton B.

APPLICANT: Koshi, Jeffrey M.

APPLICANT: Kovalic, David K.

APPLICANT: Liu, Jingdong

APPLICANT: McIninch, James

APPLICANT: Wu, Wei

TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof

FILE REFERENCE: 38-21(51237)G

CURRENT APPLICATION NUMBER: US/09/815,264

CURRENT FILING DATE: 2001-03-23

PRIOR APPLICATION NUMBER: US 09/620,392

PRIOR FILING DATE: 2000-07-19

PRIOR APPLICATION NUMBER: US 09/702,134

PRIOR FILING DATE: 2000-10-31

NUMBER OF SEQ ID NOS: 109669

SEQ ID NO 80624

LENGTH: 9648

TYPE: DNA

ORGANISM: Oryza sativa

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(9648)

OTHER INFORMATION: unsure at all n locations

US-09-815-264-80624

Query Match 2.2%; Score 47.6; DB 5; Length 9648;

Best Local Similarity 43.5%; Pred. No. 1.6;

Matches 131; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

```
QY 1010 AGAAGCGCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1069
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 5536 agcaagagagagagagagagagagagagagagagagagagagagagagagag 5595
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1070 GCGTGAATGGCGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1129
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 5596 ggcgagagagagagagagagagagagagagagagagagagagagagagagag 5655
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1130 ACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1189
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 5656 ggcgagagagagagagagagagagagagagagagagagagagagagagagag 5715
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1190 AGTGTGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1249
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 5716 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 5775
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1250 AGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1309
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 5776 tggagagagagagagagagagagagagagagagagagagagagagagagagag 5835
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1310 A 1310
```







**THIS PAGE BLANK (USPTO)**



Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com"  
 BASE COUNT 182 a 313 c 332 g 208 t 2 others  
 ORIGIN

Query Match 44.3%; Score 946.4; DB 10; Length 1037;  
 Best Local Similarity 94.5%; Pred. No. 1.4e-135;  
 Matches 1032; Conservative 2; Mismatches 3; Indels 55; Gaps 3;

QY 953 CCATCTCCGCGCGGTTCCGGGGGGAAGTGGAGAGACAGCTGTCTTACGACGAGA 1012  
 |||  
 Db 1037 CCMTCCTCCGGCGCGGTTCCGGGGGAAGTGGAGAGACAGCTGTCTTACGACGAGA 978  
 QY 1013 AGGCGCGGAGATGAGAGAGAGAGCTCATCTTCTACAGAGGGAGGTGGAGGCTGCG 1072  
 |||  
 Db 977 AGGCGCGGAGATGAGAGAGAGAGCTCATCTTCTACAGAGGGAGGTGGAGGCTGCG 918  
 QY 1073 TGGATGCGGCGCTGTGGCCACACAGATGAGACCGCGTGAAGGCCATTCCTTACCTACG 1132  
 |||  
 Db 917 TGGATGCGGCGCTGTGGCCACACAGATGAGACCGCGTGAAGGCCATTCCTTACCTACG 858  
 QY 1133 AGCAGCTGGAGCTCTTAAGATTAACCTGATGAGCTTACCACAGGTTACCCGAGT 1192  
 |||  
 Db 857 AGCAGCTGGAGCTCTTAAGATTAACCTGATGAGCTTACCACAGGTTACCCGAGT 798  
 QY 1193 CTGTGATCCAGACACTGGGCTACCTCTCTCAAGATGAGCCCTGAGGACATTCGCAAGT 1252  
 |||  
 Db 797 CTGTGATCCAGACACTGGGCTACCTCTCTCAAGATGAGCCCTGAGGACATTCGCAAGT 738  
 QY 1253 GGAATGTGACGTCCCTGAGAGACCTGGAAGGCTTGTGAAAGTGCACAAAGGGCAGGAAA 1312  
 |||  
 Db 737 GGAATGTGACGTCCCTGAGAGACCTGGAAGGCTTGTGAAAGTGCACAAAGGGCAGGAAA 678  
 QY 1313 TGAGTCTCAGAGCTCTCTGGGGGCGCCCTCCACAGGTGGCCACCTGATGACCGCTTG 1372  
 |||  
 Db 677 TGAGTCT 642  
 QY 1373 TGAAGGAGAGGGGCGAGCTAGACAAAGACACCTTACACCTTCTTACCTG 1432  
 |||  
 Db 641 TGAAGGAGAGGGGCGAGCTAGACAAAGACACCTTACACCTTCTTACCTG 582  
 QY 1433 GGTACTCTGTCT 1492  
 |||  
 Db 581 GGTACTCTGTCT 522  
 QY 1493 CGGTCAGGCGCCAGGAGCTTGAACAGCTGACCCCAAGGAGCTGAGCTCTCTATCCCA 1552  
 |||  
 Db 521 CGGTCAGGCGCCAGGAGCTTGAACAGCTGACCCCAAGGAGCTGAGCTCTCTATCCCA 462  
 QY 1553 AGGCGCGGCTTGTCTTCCAGAACATGAACGGGTCCGAATCTTCTGTAAGATCCAGTCT 1612  
 |||  
 Db 461 AGGCGCGGCTTGTCTTCCAGAACATGAACGGGTCCGAATCTTCTGTAAGATCCAGTCT 402  
 QY 1613 TCTCTGGGTGGGGGCGCCAGGAGATTTGAAGGGCTCAGTCAGAGAAATGTGACATGG 1672  
 |||  
 Db 401 TCTCTGGGTGGGGGCGCCAGGAGATTTGAAGGGCTCAGTCAGAGAAATGTGACATGG 342  
 QY 1673 ACTTGGCCACCTTATGAGCTGCGGAGAGTGGGGTCTCCGCTTGTGACTGTGGCTGAGG 1732  
 |||  
 Db 341 ACTTGGCCACCTTATGAGCTGCGGAGAGTGGGGTCTCCGCTTGTGACTGTGGCTGAGG 282  
 QY 1733 TGCAGAAACTTCTGGGAGCCCAAGTGAAGGCTGAAAGCGGAGAGAGCGGACCGCCGG 1792  
 |||  
 Db 281 TGCAGAAACTTCTGGGAGCCCAAGTGAAGGCTGAAAGCGGAGAGAGCGGACCGCCGG 222  
 QY 1793 TGCAGGAGCTGATCTTACGAGGAGCGGACGAGACCTTGGGAGCTGGGGCTGAGGCTAC 1852  
 |||  
 Db 221 TGCAGGAGCTGATCTTACGAGGAGCGGACGAGACCTTGGGAGCTGGGGCTGAGGCTAC 163  
 QY 1853 AGGCGGAGCTCCCAAGGAGCTACCGGCTCAGAGCTCAGGCTCAGAGACCTCTCTGG 1912  
 |||  
 Db 162 AGGCGGAGCTCCCAAGGAGCTACCGGCTCAGAGCTCAGGCTCAGAGCTCAGAGCTCAG 115

QY 1913 GGAGCGCCGCTCTAGAGACCTGAGACCTGTCTTACACCTGTCTGACACTGCTTACGCT 1972  
 |||  
 Db 116 -----AGGACTGAGACCTGTCTTACACCTGTCTGACACTGCTTACGCT 73  
 QY 1973 CCACCTGGCTGAGAGGCGCCAGCTCTTGTGCGCCAGCGCTGCTGGGAGATCCCGCC 2032  
 |||  
 Db 72 CCACCTGGCTGAGAGGCGCCAGCTCTTGTGCGCGCCAGCGCTGCTGGGAGATCCCGCC 13  
 QY 2033 TGGCCAGGAGCA 2044  
 |||  
 Db 12 TGGCCAGGAGTGA 1

RESULT 2  
 AL571138/c 1011 bp mRNA EST 16-FEB-2001  
 LOCUS AL571138 LTI\_NFL006.PL2 Homo sapiens cDNA clone CS0D1027YL14 3  
 DEFINITION prime, mRNA sequence.  
 ACCESSION AL571138  
 VERSION AL571138.1 GI:12928134  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 1011)  
 AUTHORS Li, W.-B., Gruber, C., Jesssee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization.  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
 Source Location/Qualifiers  
 1..1011  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1027YL14"  
 /clone.lib="LTI\_NFL006.PL2"  
 /tissue\_type="placenta"

/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
 Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com"

BASE COUNT 177 a 300 c 331 g 199 t 4 others  
 ORIGIN

Query Match 42.3%; Score 903.6; DB 10; Length 1011;  
 Best Local Similarity 96.7%; Pred. No. 4.7e-129;  
 Matches 940; Conservative 1; Mismatches 7; Indels 24; Gaps 1;

QY 1112 ACGGCATCCCTTACCTACGAGGAGCTGAGCGTCTTAACATTAATGATGAGCTCT 1171  
 |||  
 Db 948 ACGGCATCCCTTACCTACGAGGAGCTGAGCGTCTTAACATTAATGATGAGCTCT 889  
 QY 1172 ACCCAGAGTTACCCGAGCTGTGATCCAGACCTGGGCTTACTCTTCTCAAGATGA 1231  
 |||  
 Db 888 ACCCAGAGTTACCCGAGCTGTGATCCAGACCTGGGCTTACTCTTCTCAAGATGA 829  
 QY 1232 GCCCTGAGACATTCGCAAGTGAATGAGCTCCCTGGAGACCTGAAAGGCTTGTGCTG 1291  
 |||  
 Db 828 GCCCTGAGACATTCGCAAGTGAATGAGCTCCCTGGAGACCTGAAAGGCTTGTGCTG 769  
 QY 1292 AAGTCGACAAAGGCGCAAAATGATGCTCTCAGGCTCTCTGGGCGCCCTCCACAGAGTGG 1351  
 |||

```

Db 768 AAGTCACAAAGGCGACGAATGAGTCT-----CAGGTGG 733
QY 1352 CCACCTGATGACCGGCTTTGTGAAGGAAGGGCCAGCTGACAAAGACACCTAGACA 1411
Db 732 CCACCTGATGACCGGCTTTGTGAAGGAAGGGCCAGCTGACAAAGACACCTAGACA 673
QY 1412 CCGTGAACCGCTTCTACCTCTGGGTAACCTGTGCTCCCTCAGCCCGGAGAGCTGAGCTCG 1471
Db 672 CCGTGAACCGCTTCTACCTCTGGGTAACCTGTGCTCCCTCAGCCCGGAGAGCTGAGCTCG 613
QY 1472 TGGCCCCAGAGATCTGGGGCGGTCAAGGCCAGAGACCTGAGACAGTGTACCCCAAGC 1531
Db 612 TGGCCCCAGAGATCTGGGGCGGTCAAGGCCAGAGACCTGAGACAGTGTACCCCAAGC 553
QY 1532 AGCTGAGAGCTCTATCTCCCAAGGCCCGCTTGTCTTCCAGAACATGAGACGGTCCGAT 1591
Db 552 AGCTGAGAGCTCTATCTCCCAAGGCCCGCTTGTCTTCCAGAACATGAGACGGTCCGAT 493
QY 1592 ACTTGTGAAGATCAGTCTTCTGTGGGTGGGGCCCCCAGGAGATTGAAAGCGCTCA 1651
Db 492 ACTTGTGAAGATCAGTCTTCTGTGGGTGGGGCCCCCAGGAGATTGAAAGCGCTCA 433
QY 1652 GTCAAGCAATGTAGATGAGTGTGGCCAGCTGATGAGTGTGAGTGTGAGTGTGAGTGTG 1711
Db 432 GTCAAGCAATGTAGATGAGTGTGGCCAGCTGATGAGTGTGAGTGTGAGTGTGAGTGTG 373
QY 1712 TGCCGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1771
Db 372 TGCCGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 313
QY 1772 CGGAGAGCGGCGACCGCGGTGCGGGAAGTGTGATCTAGGCGAGCGGAGAGACAGCTGG 1831
Db 312 CGGAGAGCGGCGACCGCGGTGCGGGAAGTGTGATCTAGGCGAGCGGAGAGACAGCTGG 253
QY 1832 ACAGCGTGGGCGTGGGCGTACAGGGCGGCGATCCCAACAGCTGAGTGTGAGTGTGAGTGTG 1891
Db 252 ACAGCGTGGGCGTGGGCGTACAGGGCGGCGATCCCAACAGCTGAGTGTGAGTGTGAGTGTG 193
QY 1892 GCGTGAAGAGACCTCTCGGGGAGAGCGCTCTCTAGAGCTGAGAGCTGTTTCTCAACG 1951
Db 192 GCGTGAAGAGACCTCTCGGGGAGAGCGCTCTCTAGAGCTGAGAGCTGTTTCTCAACG 133
QY 1952 TCTGAGCACTGCTCTAGCTCAGCTCAGCTGAGGCGTGTGAGGCGCTGCTGAGGCGCA 2011
Db 132 TCTGAGCACTGCTCTAGCTCAGCTCAGCTGAGGCGTGTGAGGCGCTGCTGAGGCGCA 73
QY 2012 GCCCTGTGGGGATCCCGCGGTGGCCAGAGACAGAGAGGCTGATCCCGCTTCAACCCA 2071
Db 72 GCCCTGTGGGGATCCCGCGGTGGCCAGAGAGAGAGAGGCTGATCCCGCTTCAACCCA 13
QY 2072 AGAGAACTCGCG 2083
Db 12 AGAGAACTCGCG 1

```

RESULT 3  
 AL578994/c 989 bp mRNA EST 16-FEB-2001  
 LOCUS AL578994.1 L1\_NFL006.PL2 Homo sapiens cDNA clone CS0K012YE02 3  
 DEFINITION prime, mRNA sequence.  
 ACCESSION AL578994  
 VERSION AL578994.1 GI:12943602  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 989)  
 AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage

```

FEATURES
    source
        BP 191 91006 EVRY cedex - France
        Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.
        Location/Qualifiers
            1..989
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="CS0K012YE02"
                /clone_1db="L1_NFL006.PL2"
                /tissue_type="placenta"
                /note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand cDNA
                was primed with a NotI-oligo(dT) primer. Five prime end
                enriched, double-stranded cDNA was digested with Not I and
                cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                vector. Library was normalized. Library was constructed by
                Life Technologies. Contact : Feng Liang Life Technologies,
                a division of Invitrogen 9800 Medical Center Drive
                Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                Email : fliang@lifestech.com URL :
                http://fulllength.invitrogen.com"
BASE COUNT 174 a 296 c 325 g 190 t 4 others
ORIGIN
Query Match 42.0%; Score 898.6; DB 10; Length 989;
Best Local Similarity 96.1%; Pred. No. 2.8e-128;
Matches 968; Conservative 4; Mismatches 8; Indels 27; Gaps 4;
QY 1063 GAACCTGCGGTGATGCGGCGCTGCTGCGCCACCCAGATGAGACCGCGTAAGCCATCCC 1122
Db 980 GAACCTGCGGTGATGCGGCGCTGCTGCGCCACCCAGATGAGACCGCGTAAGCCATCCC 922
QY 1123 TTCACTACAGAGAGCTGAGAGCTCTAAGCATTAAGTGAAGCTTACCCAAAGT 1182
Db 921 TTCACTACAGAGAGCTGAGAGCTCTAAGCATTAAGTGAAGCTTACCCAAAGT 862
QY 1183 TACCCGAGTGTGATCAGACACCTGGGCTAAGCTTCTCAAGATGAGCCGTGAGAGC 1242
Db 861 TRCCSCGAGTGTGATCAGACACCTGGGCTAAGCTTCTCAAGATGAGCCGTGAGAGC 802
QY 1243 ATTGCGAAGTGAATGAGAGCTCCCTGAGAGACCTGAGAGGCTTGTGAAGTGAAGAAA 1302
Db 801 ATTGCGAAGTGAATGAGAGCTCCCTGAGAGACCTGAGAGGCTTGTGAAGTGAAGAAA 742
QY 1303 GGGCAGCAATGAGTCTCAGAGCTCTCGGCGGCGCTCCACAGAGTGGCCACCTGATC 1362
Db 741 GGGCAGCAATGAGTCTC-----CAGGTGGCCACCTGATC 706
QY 1363 GACCGCTTTGTGAAGGAAGGGCCAGCTAGACAAGACACCTAGACACCTGAGCGCC 1422
Db 705 GACCGCTTTGTGAAGGAAGGGCCAGCTAGACAAGACACCTAGACACCTGAGCGCC 646
QY 1423 TTCTACCTGGGTACCTGTGCTCCCTCAGCGCCGAGAGAGTGAAGTCCGTCGCCCGCAGC 1482
Db 645 TTCTACCTGGGTACCTGTGCTCCCTCAGCGCCGAGAGAGTGAAGTCCGTCGCCCGCAGC 586
QY 1483 AGCATCTGGCGGTCAAGGCCAGAGCTGAGACAGTGTGACCCCAAGGACAGCTGAGCTC 1542
Db 585 AGCATCTGGCGGTCAAGGCCAGAGCTGAGACAGTGTGACCCCAAGGACAGCTGAGCTC 526
QY 1543 CTCTATCCCAAGGCCCGCTTGTCCAGAAATGAAGGGGTCCGATCTCTGTGAG 1602
Db 525 CTCTATCCCAAGGCCCGCTTGTCCAGAAATGAAGGGGTCCGATCTCTGTGAG 466
QY 1603 ATTCAGTCTCTCTGGGTGGGGCCCCCAGAGAGATTGAAAGGGGTCAAGCAGAGAT 1662
Db 465 ATTCAGTCTCTCTGGGTGGGGCCCCCAGAGAGATTGAAAGGGGTCAAGCAGAGAT 406
QY 1663 GTGAGCATGAGACTTGGCCAGCTTATGAAAGCTGCGAGAGTGTGCGTGTGAGT 1722
Db 405 GTGAGCATGAGACTTGGCCAGCTTATGAAAGCTGCGAGAGTGTGCGTGTGAGT 346
QY 1723 GTGCTGAGGTGCAAGAACTTGTGGAGCCCGCTGAGGCGCTGAAAGGGGAGAGCGG 1782

```

	DB	345	GTTGGCTGAGTGTGCAGAAACTTCTGGGACACCACATGTGAGAGGCCCTGAAGCCGAGACGCCG	286
OY	1783	CACCGCCCGGTCGCGGACTGGAATCTCTAGGCGCAGCGGCGAGCACGACCTGGACACGCTGGGG	1842	
Db	285	CACCGCCCGGTCGCGGACTGGAATCTCTAGGCGCAGCGGCGAGCACGACCTGGACACGCTGGGG	227	
OY	1843	CTGGGGGCTACAGAGGGGGCATTCGCCAAGCGCTACGTGGTCTCTAGACCTCAAGCGGTGCAAGNG	1902	
Db	226	CTGGGGGCTACAGAGGGGGCATTCGCCAAGCGCTACGTGGTCTCTAGACCTCAAGCGGTGCAAGNG	167	
OY	1903	ACCCCTCTCGGGGAGCAGCCCTGACCTCTTAGAGACCTTGACGACCTGTCTCACCCTCTGGCATTG	1962	
Db	166	-CCCTCTCGGGGAGCAGCCCTGACCTCTTAGAGACCTTGACGACCTGTCTCACCCTCTGGCATTG	108	
OY	1963	CTCTTAAGCTCCACCTCTGAGGCCCCCACTCCTCTTGCTGGCCCCAGCCCTGCTGGG	2022	
Db	107	CTCTTAAGCTCCACCTCTGAGGCCCCCACTCCTCTTGCTGGCCCCAGCCCTGCTGGG	48	
OY	2023	GATCCCGCGCTGGCAGAGCAGCAGCGAGTATCCCGCTTCCACC	2069	
Db	47	GATCCCGCGCTGGCAGAGCAGCAGCGAGTATCCCGCTTCCACC	1	
RESULT	4			
LOCUS	AL574128	1051 bp	mRNA	EST
DEFINITION	AL574128 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0D1040TH04	5		16-FEB-2001
ACCESSION	AL574128			
VERSION	AL574128.1	GI:12934034		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 1051)			
JOURNAL	Li,W.B., Gruber,C., Jesse,J. and Polayes,D.			
COMMENT	Full-length cDNA libraries and normalization Unpublished (2001) Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr. Location/Qualifiers 1..1051 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CS0D1040TH04" /clone_1lb="LTI_NFL006.PL2" /tissue_type="Placenta" /note="Vector: pCMVSPORT 6; Site:1; NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"			
FEATURES	source			
BASE COUNT	153 a	364 C	341 G	187 T
ORIGIN				6 others

Query Match	41.7%;	Score 890.6;	DB 10;	Length 1051;
Best Local Similarity	96.1%;	Pred. No. 4.5e-127;		
Matches 997;	Conservative 2;	Mismatches 27;	Indels 12;	Gaps 8;

0y 54 TACCGGTGGACCCACAGGTGCTCCCTCCCTGGATCTACACAGACCATGGCCCTTG-CAAC 112  
|||||  
Db 18 TACCGGTGGACCCACAGGTGCTCCCTCCCTGGATCTACACAGACCATGGCCCTTGCCCAAC 77  
113 GGCCTGACCCCTGTTGGTCTGTGGGAGCCGCC-----TGGCAGCCTCCTGTTCTGCT 167

Db	78	GGCTCGACCCCTGTTGGGGTCTCTGTGGAGACCCCGCCCTCGGACGCTCTCTCTGT	137
Oy	168	CTTACAGCTTCGAATGGGTGCATCCCGGAGAGACCTGAGTGAAGACAGAGGAGAGATC	227
Db	138	CTTCAGCTCTGGATGGGTGACAGCCCTTCGAGGACCTCTGGCTGGAGACAGGCGAGAGGC	197
Oy	228	TGCCCCCTGGGGGGAGTCCCTGCAGAACCCCCCACTAACATTTCAGAGCTCTCCCTTCGCA	287
Db	198	TGCGCCCTTGAGAGGAGATCTGGGCAACCCACAC- TAACATTTCAGCTCTCTCCCTTCGCA	256
Oy	288	ACTCTTGGCTTCCCGTGTGCGAGAGTATCCGCGCTGAGACAGGAGACCTGTCCGGAGCT	347
Db	257	ACTCTTGGCTTCCCGTGTGCGAGAGTATCCGCGCTGAGACAGGAGACCTGTCCGGAGCT	316
Oy	348	GGCTGTGGCTTGGCACAGAAATGTCAAGCTCTCAACAGAGACGCTGCTGTCTGGC	407
Db	317	GGCTGTGGCTTGGCACAGAAATGTCAAGCTCTCAACAGAGACGCTGCTGTCTGGC	376
Oy	408	TCACCGGCTCTCAGAGCCCCCGAGAGACCTGCAGAGCTCTCCATTGAGACCTCTCTATT	467
Db	377	TCACCGGCTCTCAGAGCCCCCGAGAGACCTGCAGAGCTCTCCATTGAGACCTCTCTATT	436
Oy	468	CCTCAACCCAGATGCTTCTTCGGGGCCCCAGGCTCTCACCCCTTCTCTCCCGCATAC	527
Db	437	CCTCAACCCAGATGCTTCTTCGGGGCCCCAGGCTCTCACCCCTTCTCTCCCGCATAC	496
Oy	528	GAAAGCCATGTGAGACCTGCTCCCGAGAGGGGGTCTCCGAGCGACAGCG- GCTGCTGCTG	586
Db	497	GAAAGCCATGTGAGACCTGCTCCCGAGAGGGGGTCTCCGAGCGACAGCG- GCTGCTGCTG	556
Oy	587	CGGCTCTGAGCTGCTGGGGGTGTGGGGGCTCTCTG- CTAGCGAGGCTGATGTGCGGCT	645
Db	557	CGGCTCTGAGCTGCTGGGGGTGTGGGGGCTCTCTGCTAGCGAGGCTGATGTGCGGCT	616
Oy	646	CTGGAGAGCCGAGGCTTTCGACACTGCCGAGGGGCTTGTGTGGCGAGTCCGCGCACTGCTG	705
Db	617	CTGGAGAGCCGAGGCTTTCGACACTGCCGAGGGGCTTGTGTGGCGAGTCCGCGCACTGCTG	676
Oy	706	CTACCCCGGCTGTGAGCTGCCCGGAGCCCTTGAGCAGAGACAGCAGAGGAGGACGACAGG	765
Db	677	CTACCCCGGCTGTGAGCTGCCCGGAGCCCTTGAGCAGAGACAGCAGAGGAGGACGACAGG	736
Oy	766	GCGGCTCTGAGAGGGCGGGGAGACCCCTCTAGGCGCCCGCTGCGACATGCTCTGTCACAG	825
Db	737	GCGGCTCTGAGAGGGCGGGGAGACCCCTCTAGGCGCCCGCTGCGACATGCTCTGTCACAG	795
Oy	826	ATGAGACGCTTGTGCGGGGCTGTACCCCTGTCTGTGGCGAGACCCATCTCGCAGCATCCG	885
Db	796	ATGAGACGCTTGTGCGGGGCTGTACCCCTGTCTGTGGCGAGACCCATCTCGCAGCATCCG	855
Oy	886	CAGGACATCTGTGGCGCGGCTGTGGCGGCAACGCTCTCTGTGAGGACCCATCTGTGGCGAGCT	945
Db	856	CA- GCMATCTGTGGCGCGGCTGTGGCGGCAACGCTCTCTGTGAGGACCCATCTGTGGCGAGCTG	913
Oy	946	GAAAGGACATCTCTCGCGCGGCTTTCGCGCGGAGAGTGTGAGAAAGACAGCTGTCTTCA	1005
Db	914	MAAGGACATCTCTCGCGCGGCTTTCGCGCGGAGAGTGTGAGAAAGACAGCTGTCTTCA	973
Oy	1006	GGCAAGAGGCGCGGAGATAGACGAGAGCTTCATTTCTACAAAGAGTGGAGCTGGAA	1065
Db	974	GGCAAGAGGCGCGGAGATAGACGAGAGCTTCATTTCTACAAAGAGTGGAGCTGGAA	1033
Oy	1066	GCTTGGCTGATGGGCGC 1083	
Db	1034	GCTTGGCTGATGGGCGC 1051	

RESULT	5
AL556628/c	
LOCUS	
DEFINITION	AL556628 923 bp mRNA EST 16-FEB-2001 AL556628 LTL_NFLJ06_P1.2 Homo sapiens CDNA clone CSDDK006YML8 3 prime, mRNA sequence.



ACCESSION	AL556628
VERSION	AL556628.1
KEYWORDS	GI:12899477
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	L1,W.B., Gruber,C., Jesse,J. and Polayes,D.
JOURNAL	Full-length cDNA libraries and normalization
COMMENT	Unpublished (2001)
	Contact: Genoscope
	Genoscope - Centre National de Sequencage
	Bp 191 91006 EVRI cedex - France
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES	
source	Location/Qualifiers
	1..923
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="CS0DK006IN18"
	/clone_id="L1.NFL006.PL2"
	/tissue_type="placenta"
	/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT	167 a 276 c 301 g 176 t 3 others
ORIGIN	
Query Match	40.8% Score 873.2; DB 10; Length 923;
Best Local Similarity	96.5%; Pred. No. 2.1e-124;
Matches 909; Conservative	6; Mismatches 16; Indels 24; Gaps 1
Dn	1118 TCCCTTCACCTAGCAGCAGCGTGACGCTCTAAACATTAAGTGATGAGCTCAACCAC 1177
Dn	922 TCCTTTACTTAGCAGCAGCGTGACGCTCTAAACATTAAGTGATGAGCTCAACCAC 863
Dn	1178 AAGGTACCCGGAGTCTGTGATCAGCAGCTGGGCTACCTCTTCTCAAGATGAGCCCTG 1237
Dn	862 AAGGTACCCGGAGTCTGTGATCAGCAGCTGGGCTACCTCTTCTCAAGATGAGCCCTG 803
Dn	1238 AGGACATTCCGAAGTGAATGTGAGCTGCCCTGGAACCTGGAAGTTTGTGTAAGTCG 1297
Dn	802 AGGACATTCCGAAGTGAATGTGAGCTGCCCTGGAACCTGGAAGTTTGTGTAAGTCG 743
Dn	1298 ACAAAGGCACAGAATGAGTCTCTCAGGCTCCCTGGCGGCCCTCCACACAGTGGCCACCC 1357
Dn	742 ACAAAGGCACAGAATGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 707
Dn	1358 TGATGACACGGCTTTGTGAAGGGAAGGGCCAGCTAGACAAGAACACCTAGACACCTGA 1417
Dn	706 TGATGACACGGCTTTGTGAAGGGAAGGGCCAGCTAGACAAGAACACCTAGACACCTGA 647
Dn	1418 CCGCTTTACCTTGATGAGTCTGCTCCCTCAGCCCCGAGGAGCTGAGCTCCGTGCCCC 1477
Dn	646 CCGCTTTACCTTGATGAGTCTGCTCCCTCAGCCCCGAGGAGCTGAGCTCCGTGCCCC 587
Dn	1478 CCAGAGCATCGGGCGGTGAGGCCCCAGAGCCTGGACACGTTGACCCCAAGGCAAGCTGG 1537
Dn	586 CCAGAGCATCGGGCGGTGAGGCCCCAGAGCCTGGACACGTTGACCCCAAGGCAAGCTGG 527
Dn	1538 ACCTCTCTATCCCAAGGCCCGCTTGGTTTCCAACAATGAAGGAGTCCGAATACCTGG 1597
Dn	526 ACCTCTCTATCCCAAGGCCCGCTTGGTTTCCAACAATGAAGGAGTCCGAATACCTGG 467
Dn	1598 TGAAGATCCAGTCTTCTCTGGGTGGGCCCCACGAGGAGATTGGAAGGCGCTCACTGACC 1657

[illegible][illegible]





QY	2050	GGGTG	2054
Db	7	GGGTG	3
RESULT	9		
AL556862			
LOCUS	AL556862		
DEFINITION	AL556862 LTL_NFL006.PL2 Homo sapiens cdna clone CS0DK012YE02 5	EST	16-FEB-2001
ACCESSION	AL556862		
VERSION	AL556862.1	GI:12899917	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
TITLE	Li, W.B., Gruber, C., Jesses, J. and Polayes, D.		
JOURNAL	Full-length cDNA libraries and normalization		
COMMENT	Unpublished (2001)		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191.91006 EVRY cedex - France		
	Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.		
FEATURES			
Source	1..965		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="CS0DK012YE02"		
	/clone_lib="LTL_NFL006.PL2"		
	/tissue_type="placenta"		
	/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA		
	was primed with a NotI-Oligo(dT) primer. Five prime end		
	enriched, double-stranded cDNA was digested with Not I and		
	cloned into the Not I and Eco RV sites of the pCMVSPORT 6		
	vector. Library was normalized. Library was constructed by		
	Life Technologies. Contact: Peng Liang Life Technologies,		
	a division of Invitrogen 9800 Medical Center Drive		
	Rockville, Maryland 20850, USA Fax : (1) 301 610 8371		
	Email : filiang@life.com URL :		
	http://fulllength.invitrogen.com"		
BASE COUNT	135 a 340 c 314 g 171 t 5 others		
ORIGIN			
Query Match	38.1%; Score 815.4; DB 10; Length 965;		
Best Local Similarity	95.5%; Pred. No. 1.4e-115;		
Matches	920; Conservative 5; Mismatches 26; Indels 12; Gaps 8;		
QY	33	TGTGACGGCGGACAGAGACTACCGGTGAGACCAACGGTCCCTCCCTGGGATCTAC	92
Db	1	TGTGACGGCGGACAGAGACTACCGGTGAGACCAACGGTCCCTCCCTGGGATCTAC	60
QY	93	ACAGACCATGGCGTTG-CAAGCGCTCGAACCCCTGTTGGTCTGTGGGAGACGGCC- ----	146
Db	61	ACAGACCATGGCGTTGCCAACGGCTCGAACCCCTGTTGGGCTCTGTGGGACCCCGCCCT	120
QY	147	TGGCAGCCTCCGTTCTCGTCTTACGCTCGATGAGGAGCATCCCGAGGACCGCTGAC	206
Db	121	CGGACGCTCCTGTTCTGCTCTTACGCTCGGATGAGTGTGACGCTCGAGGACCGCTGAC	180
QY	207	TGAGAGACAGGAGAGAGTGTGCCCGCTGGGGGAGTCTGACAAACCCCATTAACAT	266
Db	181	TGAGAGACAGGAGAGAGGCTGCCCGCTGGACGAGAGTGTGCCAACCCAC- TAACAT	239
QY	267	TTCAGGCTCTCCCTCGGCAACTCCTTGGCTTCCCGTGTGGGAGGTGTCCGCGCTGAG	326
Db	240	TTCAGGCTCTCCCTCGGCAACTCCTTGGCTTCCCGTGTGGGAGGTGTCCGCGCTGAG	299
QY	327	CACGAGGCTGTCCGGAGAGTGTGGCTTGGACAGAGAGATGTCAACTGTCAAC	386
Db	300	CACGAGGCTGTCCGGAGAGTGTGGCTTGGACAGAGAGATGTCAACTGTCAAC	359

QY	387	ACAGACGCTGGCGCTCTGCGTCACACGGCTCTCTGAGCCCCCGACGACCGTG	-GACGGCC	445
Db	360	AGAGACGCTGGCGCTCTGCGTCACACGGCTCTCTGAGCCCCCGACGACCGTG	CGCC	419
QY	446	TCGCATTTGGACCTGCTGCTATTCCTCAACCCAGATGCTTTCGAGGAGCCACAGCCTGCA	505	
Db	420	TCGCATTTGGACCTGCTGCTATTCCTCAACCCAGATGCTTTCGAGGAGCCACAGCCTGCA	479	
QY	506	CCCGTTTCTTCTCCCGCATCACGAAGCCCATGTGACCTGCTCCGAGGGGGCTCCCG	565	
Db	480	CCCGTTTCTTCTCCCGCATCACGAAGCCCATGTGACCTGCTCCGAGGGGGCTCCCG	539	
QY	566	AGCGACAGGGGCTGCTGGCTGGCGCTGGCGCTGGCGCTGGGGGTGGGGGCTGTGGCTGA	625	
Db	540	AGCGACAGGGGCTGCTGGCTGGCGCTGGCGCTGGCGCTGGGGGTGGGGGCTGTGGCTGA	599	
QY	626	GGAGAGCTGATGTGCGGGCTGTGGAGAGCCTGGCTTGGCGACCTGCTGGGCGCTTGTGG	685	
Db	600	GGAGAGCTGATGTGCGGGCTGTGGAGAGCCTGGCTTGGCGACCTGCTGGGCGCTTGTGG	659	
QY	686	CCGAGTCGGCGCAAGTGTCTACCCCGGCTGTGTACCTGCCCGGAGCCCTGGACACAG	745	
Db	660	CCGAGTCGGCGCAAGTGTCTACCCCGGCTGTGTACCTGCCCGGAGCCCTGGACACAG	719	
QY	746	ACCACGAGGAGGACACCAAGGGGGGCTGTGCAAGGGGGGGAGACCCCTTAGGGCCCGCT	805	
Db	720	ACCACGAGGAGGACACCAAGGGGGGCTGTGCAAGGGGGGGAGACCCCTTAGGGCCCGCT	778	
QY	806	CGACATGTGTCTGTCCACGATGAGACGCTGTCCGGGGGCTGTGCTCCGCTGTGGGCGAC	865	
Db	779	CGACATGTGTCTGTCCACGATGAGACGCTGTGSGGGGCTGTGCTCCGCTGTGGGCGAC	838	
QY	866	CCATCATTCGGACGATCCCGGAGGGCATTCGTGGCGGCTGGCGGCAAGCTCTCTGGG	925	
Db	839	CCATCATTCGGACGATCCCGGA-GSCTATCTGTGGCGGCTGGG-SGCAACGCTCTCTGGG	896	
QY	926	ACCCATCTCTGGGAGGAGCTGACACGACCATCTCCGCGCCGCTTCGCGGGGAGATGG	985	
Db	897	ACCCATCTCTGGGAGGAGCTGACACGACCATCTCCGCGCGGCTTC-CGGGGAAGATGG	955	
QY	986	AGA 988		
Db	956	AGA 958		
RESULT	10			
LOCUS	AL514070	912 bp	mRNA	EST
DEFINITION	AL514070 LTI_NFL006.PL2	Homo sapiens	cdna clone	CL0BA004ZG08.5
ACCESSION	AL514070		prime, mRNA sequence.	
VERSION	AL514070.1	GI:12777564		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 912)			
JOURNAL	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.			
COMMENT	Full-length cDNA libraries and normalization			
	Unpublished (2001)			
	Contact: Genoscope			
	Genoscope - Centre National de Sequencage			
	BP 191 91006 EVRY cedex - France			
FEATURES	Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.			
source	Location/Qualifiers			
	1..912			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="CL0BA004ZG08"			
	/clone_lib="LTI_NFL006.PL2"			
	/tissue_type="placenta"			



```

|||||
Db 606 CCAAGCAGCTGAGAGCTCTCTATCCCAAGGCCCTTGCTTCCAGAACATGAACGGG 547
QY 1555 TCCGAATCTTCGTGAATATCCAGTCTCTCTGGTGGGGCCCCCAGAGAGATTGAAG 1644
Db 546 TCCGAATCTTCGTGAATATCCAGTCTCTCTGGTGGGGCCCCCAGAGAGATTGAAG 487
QY 1645 GCGCTCAGTCAAGCAATGTAGATGACTTGGCAGCTTCATGAAGCTGGCAGCGAT 1704
Db 486 GCGCTCAGTCAAGCAATGTAGATGACTTGGCAGCTTCATGAAGCTGGCAGCGAT 427
QY 1705 GCGGTCTCCCTTGACTGTGGCTGAGGTGCAAGAACTTGGGAGCCCAAGTGGAGGC 1764
Db 426 GCGGTCTCCCTTGACTGTGGCTGAGGTGCAAGAACTTGGGAGCCCAAGTGGAGGC 367
QY 1765 CTGAAGGGGAGAGAGCGCCGCGCGGCTGGGAGCTGGATCTTACGAGCGGAGAGAC 1824
Db 366 CTGAAGGGGAGAGAGCGCCGCGCGGCTGGGAGCTGGATCTTACGAGCGGAGAGAC 307
QY 1825 GACCTGAGACAGCTGGGGCTGGGGCTACAGGGCGGATCCCAACGGCTACCTGGCTTA 1884
Db 306 GACCTGAGACAGCTGGGGCTGGGGCTACAGGGCGGATCCCAACGGCTACCTGGCTTA 247
QY 1885 GACCTGAGCTGCAAGAGACCTCTCGGGGAGCGCCCTGCTCTAGACCTGGACCTGTT 1944
Db 246 GACCTGAGCTGCAAGAGAGCGCCCTCGGGGAGCGCCCTGCTCTAGACCTGGACCTGTT 187
QY 1945 CTCACCGTCTGGGACCTGCTCTAGCTCTACCTGCAACCCGCTGAGGGCCCACTCCCTTCT 2004
Db 186 CTCACCGTCTGGGACCTGCTCTAGCTCTACCTGCAACCCGCTGAGGGCCCACTCCCTTCT 127
QY 2005 GCGCCCAAGCCTGCTGGGGATCCCGCCTGGCCAGAGACAGAGGGTGAATCCCGTTTC 2064
Db 126 GCGCCCAAGCCTGCTGGGGATCCCGCCTGGCCAGAGACAGAGGGTGAATCCCGTTTC 67
QY 2065 CACCCCAAGAACTCGGCTCAGTAAACGGGAGACATCCCCCTGACAGACAAAAAAA 2124
Db 66 CACCCCAAGAACTCGGCTCAGTAAACGGGAGACATCCCCCTGACAGACAAAAAAA 7
QY 2125 AAAAAA 2130
Db 6 AAAAAA 1

RESULT 12
AL555912 817 bp mRNA EST 16-FEB-2001
LOCUS AL555912 LTI.NFL006.PL2 Homo sapiens cDNA clone CS0DK002YC06 5
DEFINITION AL555912 LTI.NFL006.PL2 Homo sapiens cDNA clone CS0DK002YC06 5
ACCESSION AL555912
VERSION AL555912.1 GI:12898014
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 817)
AUTHORS L.M.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Source
Location/Qualifiers
1..817
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK002YC06"
/clone_1bp="LTI.NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end

```

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : [filang@lifetech.com](mailto:filang@lifetech.com) URL : <http://fulllength.invitrogen.com>

BASE COUNT 109 a 285 c 270 g 151 t 2 others

ORIGIN

Query Match 31.7%; Score 677.8; DB 10; Length 817;  
Best Local Similarity 95.0%; Pred. No. 1.6e-94;  
Matches 765; Conservative 1; Mismatches 28; Indels 11; Gaps 6;

```

QY 46 CAGAGAGCTACCGGTGAGACCCAGAGTGCCTCCCTCTGGGATCTACACAGACATGGCC 105
Db 1 CAGAGAGCTACCGGTGAGACCCAGAGTGCCTCCCTCTGGGATCTACACAGACATGGCC 60
QY 106 TTG-CAAGGGCTCGACCCCTGTTGGTCTCTGTGGGAGCGGCC-----TGGCAGCTCTG 159
Db 61 TTGCAAGGGCTCGACCCCTGTTGGTCTCTGTGGGAGCGGCCCTGCGAGCTCTCTG 120
QY 160 TTCTGCTCTTCAAGCTCGATGGGTGCATCCCGCAGAGACCTGGCTGAGAGACAGAG 219
Db 121 TTCTGCTCTTCAAGCTCGATGGGTGCATCCCGCAGAGACCTGGCTGAGAGACAGAG 180
QY 220 ACGAGTGTGCCCCCTGGGGGAGTCTGACACACCCCATATACATTTCCAGCTCTTC 279
Db 181 CAGGAGCTGCGCCCTCGAGAGAGTGTGGCCACCCAC-TCACATTTCCAGCTCTTC 239
QY 280 CCTGCGCAACTCTTGGGCTTCCTCCGTGCGAGAGTGTGGGCTGAGACAGAGAGTTC 339
Db 240 CCTGCGCAACTCTTGGGCTTCCTCCGTGCGAGAGTGTGGGCTGAGACAGAGAGTTC 299
QY 340 CGGAGCTGTGGTGTGGCCCTTGGCACAGAAATGTCAAGCTCTCAACAGACAGCTGGCC 399
Db 300 CGGAGCTGTGGTGTGGCCCTTGGCACAGAAATGTCAAGCTCTCAACAGACAGCTGGCC 359
QY 400 TGTCTGGCTCACCGGCTCTGTAGACCCCGCAGAGACTGTAGAGCCCTCCATTGGACCTG 459
Db 360 TGTCTGGCTCACCGGCTCTGTAGACCCCGCAGAGACTGTAGAGCCCTCCATTGGACCTG 419
QY 460 CTGCTATTCTCAACCCAGATGCGTTTCGCGGCCCCCAGAGCTGACCCGTTTCTTC 519
Db 420 CTGCTATTCTCAACCCAGATGCGTTTCGCGGCCCCCAGAGCTGACCCGTTTCTTC 479
QY 520 CGCATCAGGAAGGCCAATGTGACCTGCTCCGAGGGGGCTCCCGAGAGCAGACGG-CT 578
Db 480 CGCATCAGGAAGGCCAATGTGACCTGCTCCGAGGGGGCTCCCGAGAGCAGACGGGCT 539
QY 579 GCTGCTCGCGCTGTGGCTGTGGGGTGTGGGGGTCTCTGCTGAGCGAGCTGATGT 638
Db 540 GCTGCTCGCGCTGTGGCTGTGGGGTGTGGGGGTCTCTGCTGAGCGAGCTGATGT 599
QY 639 GCGGGCTGTGGAGAGCCGTGGCTGGAGCTGCTGGGGGCTTTGGGCGAGTGGGCGGA 698
Db 600 GCGGGCTGTGGAGAGCCGTGGCTGGAGCTGCTGGGGGCTTTGGGCGAGTGGGCGGA 659
QY 699 AGTGTGCTACCCCGGCTGTGAGCTGCCCGGAGCCCTTGAGACAGACAGCAGAGAGGC 758
Db 660 AGTGTGCTACCCCGGCTGTGAGCTGCCCGGAGCCCTTGAGACAGACAGCAGAGAGGC 719
QY 759 AGCCAGGGCGGCTGTGAGGGCGGGGAGACCCCTACAGGCCCTCCGCTGACATGTCTGT 818
Db 720 AGCCAGGGCGGCTGTGAGGGCGGGGAGACCCCTACAGGCCCTCCGCTGACATGTCTGT 777
QY 819 CTCACAGATGAGCGCTGCGGGGC 843
Db 778 CTCACAGATGAGCGCTGCGGGGC 801

```

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	JOURNAL COMMENT	FEATURES SOURCE
AI813749/c	AI813749 wk79a05.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2421584 3' similar to TF:Q14859 Q14859 PRE-PRO-MEGAKARYOCYTE POTENTIALTING FACTOR PRECURSOR. ; , mRNA sequence.	AI813749	AI813749.1	GI:5424964	EST. human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	Contact: Robert Strausberg, Ph.D. Email: cgabds.remali.nih.gov Life Technologies catalog # : 11548-013 DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <a href="http://www.bio.lnl.gov/dbrr/image/image.html">www.bio.lnl.gov/dbrr/image/image.html</a> Insert Length: 2146 Std Error: 0.00 Seq primer: -400P from Glibco High quality sequence stop: 424.	1. 737 location/Qualifiers organism="Homo sapiens" db_xref="taxon:9606" clone_image="IMAGE:2421584" clone_lib="NCI_CGAP_Pan1" tissue_type="adenocarcinoma" lab_host="DH10B" note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: Salt; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"
BASE COUNT	126 a 230 c 241 g 139 t 1 others						
ORIGIN							
Query Match	31.7%; Score 677; DB 10; Length 737;						
Best Local Similarity	95.7%; Pred. No. 2.2e-94;						
Matches 706; Conservative	0; Mismatches 31; Indels 1; Gaps 1;						
OY 1377	GCGAAGGGGCCACTAGACAAGAACCCTTAGACACCTGCCTTCCTACCTGGGTGA 1436	                   					
Db 737	GCGCAGGGGGCAAGCTAGA-AAAGAACACTTAAGATCCCTCGAGGCTTTTAACTTGCGGTGA 679	                   					
OY 1437	CCTGTGTCCTCCACACCCCGAGAGACTGTAGTCGTCGTGGCCCCCACAGCATCTGGGGCGGT 1496	                   					
Db 678	CCTGTGTCTCCTCACACACAGAAACCTGTAGCTCCGTGCCCCAGCTTAGCATTTGGGGGT 619	                   					
OY 1497	CAGGCCCCAGAGACTCGACACCTGTGTAGCCCAAAGCAGCTGTGACGTCCTATCCCAAGGC 1556	                   					
Db 618	CAGGCCCCAGAGACTGTGACACAGTGTGACCCAAAGCAGCTGTGACGTCCTATCCCAAGGC 559	                   					
OY 1557	CCGCTTCCTTCCAGAACATGACAGCGGTCCGAACTCTTGTAAGATCCAGTCTTCT 1616	                   					
Db 558	CCGCTTCCTTCCAGAACATGACAGCGGTCCGAACTCTTGTAAGATCCAGTCTTCT 499	                   					
OY 1617	GGGTGGGGGGCCCCCAGAGAGATTGAAGCGCTCAGTACAGCAAGANMGAGATGGAGCT 1676	                   					
Db 498	GGGTAGGGCCCCCAGAGAGATTGAAGCGCTCAGTACAGCAAGATGTAGATGGACTT 439	                   					
OY 1677	GCGCAGTTCATGAGAGCTGCGAGACGATGCGCTGCTGCGCTTGACTGTGGCTGAGAGTGA 1736	                   					
Db 438	GCGCAGTTCATGAGAGCTGCGAGACGATGCGCTGCTGCGCTTGACTGTGGCTGAGAGTGA 379	                   					
OY 1737	GAACTTCTGGGACCCCACGTCGAGGGCTTGAAGCGGAGAGAGCGGACCGCCCGGTGCG 1796	                   					

Dd	378	GAACTTGTGGGACCCCACTGGAGGSCCTGAAGCGGAGAGCGGCACGCCCGGTGCG	319
Qy	1797	GGACTGATCTCTACGGCAGCGGACGACGACCTGTGACACGCTGGGGCTTGAGGCTACAGGG	1856
Dd	318	GGACTGGATCTCTACGGCAGCGGACGACGACCTGTGACACGCTGGGGCTTGAGGCTACAGGG	259
Qy	1857	CGGCATCCCCCAAGCGCTACCTGGTCTCTAGACTCTAGCGCTGCAGAAGAACCTCTTGSGGGAC	1916
Dd	258	CGGCATCCCCCAAGCGCTACCTGGTCTCTAGACTCTAGCATGTGAMAAGGCCCTCTGSGGGAC	199
Qy	1917	GCCCTGACCTCTAGGACCTGGACCTGTTCTCFACCGCTCTGTGGACACTGCTCTAGACTCCAC	1976
Dd	198	GCCCTGACCTCTAGGACCTGGACCTGTTCTCFACCGCTCTGTGGACACTGCTCTAGACTCCAC	139
Qy	1977	CCTGGCTTGAGGGGCCCCCACTCCCTGTGCTGGGCCCCAGCCCCTCTGTGGGGATTCCCGCTGGC	2036
Dd	138	CCTGGCTTGAGGGGCCCCCACTCCCTGTGCTGGGCCCCAGCCCCTCTGTGGGGATTCCCGCTGGC	79
Qy	2037	CAGGACGAGGACGAGGGGTATCCCGCTTCACCACCCAGAGMAACTCGCGCTCAGTAACGGG	2096
Dd	78	CAGGACGAGGACGAGGGGTATCCCGCTTCACCACCCAGAGMAACTCGCGCTCAGTAACGGG	19
Qy	2097	AACATGCCCCCTGCAGAC	2114
Dd	18	AACATGCCCCCTGCAGAC	1
RESULT	14		
LOCUS	AI984029/c		
DEFINITION	wt32e09.x1 NCI-CGAP_PanI Homo sapiens cDNA clone IMAGE:251112 3'	EST	09-MAR-2000
	similar to TR:Q14859 Q14859 PRE-PRO-MEGAKARYOCYTE POTENTIATING		
	FACTOR PRECURSOR.; contains MSRI.t2 MSRI repetitive element ; , mRNA		
SEQUENCE.			
ACCESSION	AI984029		
VERSION	AI984029.1	GI:5811248	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 759)		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapds@mail.nih.gov Life Technologies catalog #: 11548-013 DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.lnl.gov/dbrr/image/image.html Insert Length: 1518 Std Error: 0.00 Seq primer: -40UP from Glibco High quality sequence stop: 419.		
FEATURES			
Source	Location/Qualifiers		
	1..759		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_image="IMAGE:251112"		
	/clone_lib="NCI-CGAP_PanI"		
	/tissue_type="adenocarcinoma"		
	/lab_host="DH10B"		
	/note="Organ: pancreas; Vector: pCMV-Sport6; Site_1: Salt; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.72 kb. Life Technologies catalog #:		
	11548-013"		
BASE COUNT	131 a 237 c 232 g 155 t	4 others	
ORIGIN			



Query Match	30.4%	Score: 650.2	DB: 10	Length: 759
Best Local Similarity	93.4%	Pred. No. 2.7e-90		
Matches: 709	Conservative	0	Mismatches: 47	Indels: 3
QY 1366	CGCTTTGTGTAAGGGAAGGGCCAGCTAGACAAGACACCTTGAGACACCTGACCGCTTC	1425		
Db 757	CGCTTTGTGTAAGGGAAGGGCCAGGTAAGC-AAGACACTTAAGACACTTGGCGGCTTTN	699		
QY 1426	TACCTTGGTACCTGTGCTCCCTCAGACCCCGAGAGACTGAGTCCGTGCCCGCCAGACG	1485		
Db 698	TA-CMTGGGTACTCTGGGCTCCTCAGGGCCGAGGAAGTTAGTCGGGTGCCCGCAAGCAGC	640		
QY 1486	ATCTGGGCGGTGAGGCGCCAGAGACGTTGGACAGCTGTGAGCCAGAGGAGCTGGACGTC	1545		
Db 639	ATTTGGGGGTGATAGGCGCCAGACTGGACAGTGTGACNCAAGGAGTGGAAAGCTTTC	580		
QY 1546	TATCCCAAGGCCCGGCTTGTCCAGAACATGAACGGGTCCGAATACTTCTGTAAGATC	1605		
Db 579	TATCCCAAGGCCCGGCTTGTCCAGAACATGAACGGGTCCGAATACTTCTGTAAGATC	520		
QY 1606	CAGTCTTCTCTGGGTGGGGCCCGCCAGAGAGATTGAAAGCGCTCAGTACAGAAATGTG	1665		
Db 519	CAGTCTTCTCTGGGTGGGGCCCGCCAGAGAGATTGAAAGCGCTCAGTACAGAAATGTG	460		
QY 1666	AGCAATGAGACTTGGCCAGCTTCATGAAAGCTGGCGAGAGATGGGTGCGTGGTACTGTG	1725		
Db 459	AGCAATGAGACTTGGCCAGCTTCATGAAAGCTGGCGAGAGATGGGTGCGTGGTACTGTG	400		
QY 1726	GCTGAGGTGACAAACTTCTGTGGACCCACAGTGAAGGGCTTGAAGCGGAGAGCGGCAC	1785		
Db 399	GCTGAGGTGACAAACTTCTGTGGACCCACAGTGAAGGGCTTGAAGCGGAGAGCGGCAC	340		
QY 1786	CGCCCGGTGGGGGACTGATCTTACGGCAGCGGACAGGAGACTTGGACAGCTGGGGCTG	1845		
Db 339	CGCCCGGTGGGGGACTGATCTTACGGCAGCGGACAGGAGACTTGGACAGCTGGGGCTG	280		
QY 1846	GGGCTACAGAGCGGGGATCCCAAGGGCTACGTGACCTGAGCCTGAGCTGCAAGAGAC	1905		
Db 279	GGGCTACAGAGCGGGGATCCCAAGGGCTACGTGACCTGAGCCTGAGCTGCAAGAGAC	220		
QY 1906	CTCTC-GGGGAGCGCCTGCTCTAGAGACTTGGACCTGTTCTACCGTCTGGCACTGCT	1964		
Db 219	CTTGTGGGGGAGCGCCTGCTCTAGAGACTTGGACCTGTTCTACCGTCTGGCACTGCT	160		
QY 1965	CCTACCTCCACCCGCGCTGAGGGCCCGCACCTCTGTGTGGCCCGACGCTGTGGGGA	2024		
Db 159	CCTACCTCCACCCGCGCTGAGGGCCCGCACCTCTGTGTGGCCCGACGCTGTGGGGA	100		
QY 2025	TCCCGCGCTGGGACAGAGACAGGCGGGGTATCCCGTCCACCCCAAGAGAACTCGCGC	2084		
Db 99	TCCCGCGCTGGGACAGAGACAGGCGGGGTATCCCGTCCACCCCAAGAGAACTCGCGC	40		
QY 2085	TCAGTAAACGGGACATGCGCCCTCGACAGCAAAAAA	2123		
Db 39	TCAGTAAACGGGACATGCGCCCTCGACAGCAAAAAA	1		
RESULT 15				
LOCUS	AI744747/c			
DEFINITION	tr14b1..x1 NCI CGAP Ov23 Homo sapiens cDNA clone IMAGE:2218769 3'			
KEYWORDS	similar to TP:Q14855 Q14859 PBE-PRO-MEGAKARYOCYTE POTENTIATING			
SOURCE	FACTORY PRECURSOR.; contains MSRL.C2 MSRL repetitive element ; , mRNA			
ACCESSION	AI744747			
VERSION	AI744747.1			
KEYWORDS	EST.			
ORGANISM	human.			
REFERENCE	1 (bases 1 to 799)			
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.			

FEATURES	source	JOURNAL COMMENT	TITLE
133 a 244 c 266 g 150 t 6 others	/Organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2218269" /clone_idb="NCI_CGAP_Ov23" /tissue_type="tumor, 5 pooled (see description)" /lab_host="DH10B" /note="Organ: Ovary; Vector: PCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.35 kb. Tumor types include: mixed Mullerian tumor, papillary serous, clear cell, spindle cell. All are primary tumors, metastasis positive. Life Technologies catalog #: 11534-013"	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgabs-r@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Greg Lennon, Ph.D. Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://www.bio.llnl.gov/bbrp/image/image.html">www.bio.llnl.gov/bbrp/image/image.html</a> Seq primer: -400p from Gibco High quality sequence stop: 403.	Location/Qualifiers 1..799

Query Match	Similarity	30.1%	Score 64.8	DB 10	Length 799
Best Local	Similarity	94.0%	Pred. No.2.5e-89		
Matches 721	Conservative	0	Mismatches 41	Indels	Gaps
OY	1351	GCCACCCCTGATGACCGCGCTTTGTGAAGGGAAGGGGCGACCTGTGACAAACACACCTTAGAC	1410		
Db	767	GCCCMCTGATGCGCCGCGTTGTAAAGGGAAGGGGCCAAGTAGACAAAGCCTTAGAC	708		
OY	1411	ACCTGACCGCCTT-CYACCTGGGTAACCTGTG-CTCCCTCAGCCCCG-AGGAGCTGAGC	1467		
Db	707	ACCTGACCGCTTTTNCCTACCTGGGGGTACTGTGCTCCCTCCACGCCACGNAAGGAGCTGAGC	648		
OY	1468	TCCGGGCCCCCAGAGAGCATCTGGGGGGGTCAAGCCCCAGAGCCTGGACACGTTGACCCA	1527		
Db	647	TCCGGGCCCCCAGAGAGCATCTGGGGGGGTCAAGCCCCAGAGCCTGGACACGTTGACCCA	588		
OY	1528	AGGCAGCTGGAGCGTCCCTCATCCCAAGGCGCCGCTTGC-TTTCAGAACATGAAACGGTTC	1586		
Db	587	GGGCAAGCTGGAGGTCTCCATCCCAAGGCGCCGCTTGTTCAGAACATGAAACGGTTC	528		
OY	1587	CGAATACTTCTGGAAGATCCAGTCCCTCTCGGTGGGGCCCCACGGAGATTTGAAGCC	1646		
Db	527	CGAATACTTCTGGAAGATCCAGTCCCTCTCGGTGGGGCCCCACGGAGATTTGAAGCC	468		
OY	1647	GCTCAGTAGCAGAAATGTGAGCATGAGTCTGGCCACGTCATGAAGCTCGGAGCGGATGC	1706		
Db	467	GCTCAGTAGCAGAAATGTGAGCATGAGTCTGGCCACGTCATGAAGCTCGGAGCGGATGC	408		
OY	1707	GGTGTGCGCTTGATCTGTGGCTGAGGTGCAGAACTTCTGGGACCCCACTGTGAGGGCCT	1766		
Db	407	GGTGTGCGCTTGATCTGTGGCTGAGGTGCAGAACTTCTGGGACCCCACTGTGAGGGCCT	348		
OY	1767	GAAGCGGAGAGCGGACCGCCGCGTGCAGGACTGTGATCTTAAGGACGGCAGGACGCA	1826		
Db	347	GAAGCGGAGAGCGGACCGCCGCGTGCAGGACTGTGATCTTAAGGACGGCAGGACGCA	288		
OY	1827	CCTGACACACGCTGGGGGCTGAGGCTACAGGGGGCGCATCCCAACGGCTACTGTGCTTAA	1886		
Db	287	CCTGACACACGCTGGGGGCTGAGGCTACAGGGGGCGCATCCCAACGGCTACTGTGCTTAA	228		

```

QY 1887 CCTCAGCGTGCAGAGACCTCTCGGGGACG-CCCTGCGCTCTAGAGACCTGGACCTGTTT 1945
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 227 CCTCAGCATGCAGAGAGGCCCTCTGGGGACGCCCTGCTCTCTAGAGACTGGACCTGTTT 168
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1946 TCACCGTCTGGGCACTGCTCTAGCCTTCACACCTGGCCCTGAAGGCCCCACTTCCTTGCTG 2005
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 167 TCACCGTCTGGGCACTGCTCTAGCCTTCACACCTGGCCCTGAAGGCCCCACTTCCTTGCTG 108
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2006 GCGCCAGCGCTGCTGGGGATTCGCCGCTGGCCAGAGAGGACAGGGGTGATCCCGTTCC 2065
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107 GCGCCAGCGCTGCTGGGGATTCGCCGCTGGCCAGAGAGGACAGGGGTGATCCCGTTCC 48
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2066 ACCCAAGAGAACTCGCGCTCAGTAAACGGGGAACATGCCCGCTGCAG 2112
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47 ACCCAAGAGAACTCGCGCTCAGTAAACGGGGAACATGCCCGCTGCAG 1
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: December 7, 2001, 11:43:17  
 Job time: 6035 sec

**THIS PAGE BLANK (uspto)**



XX	Chang K, Pastan I;
PI	
PI	WPI; 1997-372620/34.
DR	
XX	N-PSDB; AAT91079.
XX	
XX	Agents for targeting mesothelin, a tumour cell antigen - used for
PT	the detection or inhibition of growth of e.g. mesotheliomas, ovarian
PT	cancers and squamous cell carcinoma
XX	
PS	Claim 1; Page 58-60; 72pp; English.
XX	
XX	This protein comprises the human CAK1 antigen which is found
CC	on mesothelium, mesotheliomas, ovarian cancers and some
CC	squamous cell carcinomas. The antigen has been designated
CC	mesothelin. Its amino acid sequence was deduced from a cDNA
CC	clone (see AAT91079) isolated from a HeLa cDNA library. A 40 kDa
CC	form (K1) of mesothelin that is found on the surface of human
CC	ovarian tumour OVCAR-3 cells appears to be derived from the 69
CC	kDa precursor by several processing steps. A claimed method for
CC	specifically delivering an effector molecule to a tumour cell
CC	bearing an antigen comprising at least 10 contiguous amino acids
CC	of mesothelin involves: (a) providing a chimeric molecule comprising
CC	the effector molecule attached to a targeting molecule that
CC	specifically binds to mesothelin; and (b) contacting the tumour with
CC	the chimeric molecule such that the chimeric molecule specifically
CC	binds to a tumour cell. Also claimed is a method for inhibiting
CC	mesothelin expression or activity by contacting mesothelin bearing
CC	cells with inhibitory nucleic acids for the mesothelin gene. The
CC	methods can be used to detect tumour cells and to inhibit the
CC	growth of cells bearing mesothelin. Mesothelin-derived antigens
CC	may be used in vaccines for the inhibition or prevention of
CC	mesotheliomas or ovarian tumours.
XX	
XX	Sequence 628 AA;

Query Match	Similarity	Score	DB	Length
Best Local	Similarity	100.0%	3261	18
Matches	628	Conservative	0	Mismatches
			0	Indels
			0	Gaps
QY 1	1	1	1	1
Db 1	1	1	1	1
QY 61	61	61	61	61
Db 61	61	61	61	61
QY 121	121	121	121	121
Db 121	121	121	121	121
QY 181	181	181	181	181
Db 181	181	181	181	181
QY 241	241	241	241	241
Db 241	241	241	241	241
QY 301	301	301	301	301
Db 301	301	301	301	301
QY 361	361	361	361	361
Db 361	361	361	361	361
QY 421	421	421	421	421
Db 421	421	421	421	421
QY 481	481	481	481	481
Db 481	481	481	481	481

Db 461 vlypaxaiaiafmgmgseyfvaixqsfliqaparedikaalsqavmsndlatmktldavip1 540

Qy 541 TVAEYOKLLHGHVEGLKAEERHREVRDWILRQRODDDLTGLGSGIPNGYLVLDLSVQ 600

Db 541 tvaeyokllhghveglkaeeerhrrpvrldwllrqrdldidltlglsfgqgipngylvldlsvq 600

Qy 601 ETLSTGCLLGGPVLTVALLLSTLA 628

Db 601 etlstgpclllgpvtlvalllslsla 628

RESULT	2
AAR53992	
ID	AAR53992 standard; Protein; 622 AA

AC AAR53992;

DT 08-DEC-1994 (first entry)

DE Megakaryocyte potentiator.

KW Megakaryocyte potentiator; Meg-Pot; thrombocytopenia,  
KM platelet.

XX Homo sapiens

XX	Key	Location/Qualifiers
FH		

FT	Protein	39.0622
FT		/note="claim 1"

FI	FOREIGN	57,1400
FT		/note= "claim 6"

PN W09410312-A.

PD 11-MAY-1994

PF 25-OCT-1993; 93WO-JP01540.

PR	23-OCT-1992;	92JP-0286153.
PR	11-NOV-1992;	92JP-0301387.

PR 09-DEC-1992; 92JP-0329546.  
XX

PA (CHUS ) CHUGAI SEIYAKU KK.  
XX

Pl Hattori K, Kojima T  
XX

DR WPL; 1994-10/401/20.  
DR N-PSDB; AAQ63972.

PT New mega:karyoctye potentiator - for potential treatment of

XXXXXXXXXXXX

[illegible]

CC use in treatment of thrombocytopenia and low platelet function

**SQ** Sequence 622 AA;

Query Match	93.6%;	Score 3053;	DB 15;	Length 622;
Best Local Similarity	95.4%;	Pred. No. 2.4e-270;		
Matches 601;	Conservative	3;	Mismatches 16;	Indels 10; Gaps 3

Oy	MAORLQD-CWSCDRP-GSLTFLFSLGAWHPARTLAGEGTSSAPLDGLVTPHHS	58
Dd	1 malptarpllyscgtpnagslfllflfsgwqpsrtlaetggcaaplldylanpniss	60
Oy	LSPRLQLGFCAEYSGSLSTERYRELAAVALAOKNKLSTEDLRCLAHRLSEPPEDDALPL	118
Dd	61 lspqlrlgfcaevsygstlervtelavalaknvkstegrlrcahlisepeddalpi	120

```

OY 119 DLLFLNPDAFSGPOACTREFSRITKANVDLPRGAPERORLLPAALACWGVRSLLSEA 178
DB 121 dlilflnpdaispgpactrefsritkanvdlprgapergrllpaalacwgvrsllsea 180
OY 179 DVRAAGLACDLPGRVAESAELVLPRLVSCPGPLDQDQGAARALQGGGPPYGPSTW 238
DB 181 dvralgglacdlpgrfvaesaevllprlvscppldqdggeaaraa1999ppypgstw 240
OY 239 SVSTMALRLGLVLTGPIRSTPGCTVAAWRQRSSRDPGWRQPERITLPRPRREVEKT 298
DB 241 svstmallrlglpvlrgpirlrstpgctvaawrqrssrdpsrqpertllprprrevekt 300
OY 299 ACPSGKKAREIDSLIFKKWELEACVDAALLATQMDRVNAIPFTYEQLDKHKLDEL 358
DB 301 acpsgkkareidesliffykkweleacvdaallatqmdrvnaipftyeqdlvkhklidely 360
OY 359 POGYPSVIOHGLYFLKMSPEDIRKKNVTSLETALKALLEVDKHEMSPOAPRRPLQVA 418
DB 361 pogypsvighlgyflfkmspedirknvtsletalkallevnykhems-----pqva 412
OY 419 TLIDRFYKRGOLDKDTLDTLTAFFPGYLCSLSPBELSVPPSSITAVRPODDLTCDPRQ 478
DB 413 tldirfkyrgldkdtldtltatfpygylcslspeelsvppssitavrpoddltcdprq 472
OY 479 LDVLYPKARLAFQNMNGSEYFVKIQSFLGAPTEDELKALSOQNVSMDLATFMKLRTDAVL 538
DB 473 ldvlypkarlafqnmngseyfvkifqslgapedelkalsqnvsmldatfmklrtdav1 532
OY 539 PLTYAEVQKILGPVHEGLKAEBHRHPRVDMILRQRODDLTGLGLOGGIPNGLYVDLS 598
DB 533 pltyaevqkllgpvhveglkaeerhprvdmilrqroddltglgloggipnglyvdls 592
OY 599 VOETLSTPCLLPGPVLTVALLASTLA 628
DB 593 vteglstpcllpgpvlvtvallaastla 622

RESULT 3
AAB08544
ID AAB08544 standard; Protein; 328 AA.
AC AAB08544;
XX
XX 20-DEC-2000 (first entry)
DE Mesothelin related antigen (MRA)-2 polypeptide sequence.
XX
XX Mesothelin; mesothelin related antigen; MRA; malignant condition; cancer;
KW differentiation antigen; adenocarcinoma; mesothelioma; ovarian carcinoma;
XX pancreatic carcinoma; non-small cell lung carcinoma; MRA-2.
XX
XX Homo sapiens.
OS
XX
XX WO200050900-A2.
PN
XX
XX 31-AUG-2000.
PD
XX
XX 25-FEB-2000; 2000MO-USO4834.
PE
XX
XX 26-FEB-1999; 99US-0121767.
PR 05-AUG-1999; 99US-0147404.
XX
XX (PACT-) PACIFIC NORTHWEST RES FOUND.
PA
XX
XX Scholler NB, Hellstrom I, Hellstrom KE;
XX
XX MPI; 2000-572118/53.
DR N-PSDB; AAA64363.
XX
XX Diagnosing carcinomas, e.g. adenocarcinoma or ovarian carcinomas, by
PT detecting reactivity of a molecule (with an antigenic determinant and
PT present in a sample) with an antibody specific for a mesothelin related
PT antigen polypeptide

```

```

XX
XX Claim 48; Fig 6A-B; 92pp; English.
PS
XX
CC The present sequence represents a human mesothelin related antigen
CC (MRA)-2. Antibodies specific to MRA polypeptides are used for the
CC detection of a malignant condition. Mesothelin is a differentiation
CC antigen which is expressed in the surfaces of normal mesothelial cells
CC and also on certain cancer cells, including epithelial ovarian tumours
CC and mesotheliomas. The method is useful for detecting the presence of a
CC malignant condition, specifically, adenocarcinoma, mesothelioma,
CC ovarian carcinoma, pancreatic carcinoma or non-small cell lung carcinoma.
XX
XX Sequence 328 AA:
SQ
Query Match 47.4%; Score 1547; DB 21; Length 328;
Best Local Similarity 97.1%; Pred. No. 5,5e-133;
Matches 301; Conservative 1; Mismatches 0; Indels 8; Gaps 1;
OY 291 FRRVEKTAQPSGKKAREIDSLIFKKWELEACVDAALLATQMDRVNAIPFTYEQLDVL 350
DB 1 frrevetlaqpsgkkareidesliffykkweleacvdaallatqmdrvnaipftyeqdlv1 60
OY 351 KHKLDELTPGYPSVIOHGLYFLKMSPEDIRKKNVTSLETALKALLEVDKHEMSPOAP 410
DB 61 khklidelypgypsvighlgyflfkmspedirknvtsletalkallevnykhems---- 116
OY 411 RRPLPOVATLIDRFYKRGOLDKDTLDTLTAFFPGYLCSLSPBELSVPPSSITAVRPOD 470
DB 117 ---pqvatlldrtfvkrgldkdtldtltatfpygylcslspeelsvppssitavrpod 172
OY 471 LDTCDPRLDVLVYPKARLAFQNMNGSEYFVKIQSFLGAPTEDELKALSOQNVSMDLATFM 530
DB 173 ldtcdprqldvlypkarlafqnmngseyfvkifqslgapedelkalsqnvsmldatfm 232
OY 531 KLRTDAVLPLTYAEVQKILGPVHEGLKAEBHRHPRVDMILRQRODDLTGLGLOGGIPN 590
DB 233 klrtdav1pltyaevqkllgpvhveglkaeerhprvdmilrqroddltglgloggipn 292
OY 591 GYIVLDLSVQ 600
DB 293 gyivldlsvq 302

RESULT 4
AAB08547
ID AAB08547 standard; Protein; 402 AA.
AC AAB08547;
XX
XX 20-DEC-2000 (first entry)
DE Soluble mesothelin related (SMR) antigen polypeptide sequence.
XX
XX Mesothelin; mesothelin related antigen; MRA; malignant condition; cancer;
KW differentiation antigen; adenocarcinoma; mesothelioma; ovarian carcinoma;
XX pancreatic carcinoma; non-small cell lung carcinoma; MRA-2; SMR antigen.
XX
XX Homo sapiens.
OS
XX
XX WO200050900-A2.
PN
XX
XX 31-AUG-2000.
PD
XX
XX 25-FEB-2000; 2000MO-USO4834.
PE
XX
XX 26-FEB-1999; 99US-0121767.
PR 05-AUG-1999; 99US-0147404.
XX
XX (PACT-) PACIFIC NORTHWEST RES FOUND.
PA
XX
XX Scholler NB, Hellstrom I, Hellstrom KE;
XX
XX
XX

```

DR WPI: 2000-572118/53.  
 DR N-PSDB: AAA64366.  
 PT Diagnosing carcinomas, e.g. adenocarcinoma or ovarian carcinomas, by  
 PT detecting reactivity of a molecule (with an antigenic determinant and  
 PT present in a sample) with an antibody specific for a mesothelin related  
 PT antigen polypeptide  
 XX  
 PS Claim 48; Fig 7A-B; 92pp; English.  
 XX  
 CC The present sequence represents a human soluble mesothelin related (SMR)  
 CC antigen. Antibodies specific to mesothelin related antigen (MRA)  
 CC polypeptides are used for the detection of a malignant condition.  
 CC Mesothelin is a differentiation antigen which is expressed in the  
 CC surfaces of normal mesothelial cells and also on certain cancer cells,  
 CC including epithelial ovarian tumours and mesotheliomas. The method is  
 CC useful for detecting the presence of a malignant condition, specifically,  
 CC adenocarcinoma, mesothelioma, ovarian carcinoma, pancreatic carcinoma  
 CC or non-small cell lung carcinoma.  
 CC  
 CC Sequence 402 AA:  
 SQ  
 Query Match 47.4%; Score 1547; DB 21; Length 402;  
 Best Local Similarity 97.1%; Pred. No. 7.5e-133;  
 Matches 301; Conservative 1; Mismatches 0; Indels 8; Gaps 1;  
 OY 291 FRREKTRACPSGKKAREIDSLIFKKWELEACVDALATQMDRVNAIPFTYEQDLYL 350  
 DB 1 firevektacpsgkkaareidesliffkkweleacvdaallatqmdrvnaipfyeqldv1 60  
 OY 351 KHKLEDELPGYPSVSIQHLGYLFLKMSPEDIRKWNVTSLETALKALLEVDKGHEMSPOAP 410  
 DB 61 khkldelypgypesvsiqhlgyilfkmspedirkwnvtsletalkallevnkghems---- 116  
 OY 411 RRPPLQVATLIDRFVKGKGGQLDKDLDLTLTAFYPGYLCSSPEELSSVPSSIAVVRPQD 470  
 DB 117 ---pqvatllidrfvkggrgldkldtltltafypgylcsspeelssvpsaiwvrpqd 172  
 OY 471 LDTCPROLDVLYPKARLAFQNMNGSEYFVKIQSFLGAPREDIKALSOQNVSMDLATFM 530  
 DB 173 ldtcprqldvlypkarlaifqnmngseyfvkigsflggapredikalsgqnvsmldatfm 232  
 OY 531 KLRDVALPLTVAEVQKLLGPHVEGLKAEERHRRPVDMILRORODDLTLGLGOGGIPN 590  
 DB 233 klrtdavplltvaevqkllgphveglkaeerhrpvrdmllrqrddldtlglg9gipn 292  
 OY 591 GYLVLDLSVQ 600  
 DB 293 gylvldlsvq 302  
 RESULT 5  
 ID AAB08543 standard; Protein: 399 AA.  
 AC AAB08543:  
 XX  
 DT 20-DEC-2000 (first entry)  
 XX  
 DE Mesothelin related antigen (MRA)-1 polypeptide sequence.  
 XX  
 KW Mesothelin; mesothelin related antigen; MRA; malignant condition; cancer;  
 KW differentiation antigen; adenocarcinoma; mesothelioma; ovarian carcinoma;  
 KW pancreatic carcinoma; non-small cell lung carcinoma; MRA-1.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200050900-A2.  
 XX  
 PD 31-AUG-2000.  
 XX  
 PF 25-FEB-2000; 2000WO-US04834.

XX  
 PR 26-FEB-1999; 99US-0121767.  
 PR 05-AUG-1999; 99US-0147404.  
 XX  
 PA (PACI-) PACIFIC NORTHWEST RES FOUND.  
 PI Scholler NB, Hellstrom I, Hellstrom KE;  
 XX  
 DR WPI: 2000-572118/53.  
 DR N-PSDB: AAA64362.  
 PT Diagnosing carcinomas, e.g. adenocarcinoma or ovarian carcinomas, by  
 PT detecting reactivity of a molecule (with an antigenic determinant and  
 PT present in a sample) with an antibody specific for a mesothelin related  
 PT antigen polypeptide  
 XX  
 PS Claim 48; Fig 5A-B; 92pp; English.  
 XX  
 CC The present sequence represents a human mesothelin related antigen  
 CC (MRA)-1. Antibodies specific to MRA polypeptides are used for the  
 CC detection of a malignant condition. Mesothelin is a differentiation  
 CC antigen which is expressed in the surfaces of normal mesothelial cells  
 CC and also on certain cancer cells, including epithelial ovarian tumours  
 CC and mesotheliomas. The method is useful for detecting the presence of a  
 CC malignant condition, specifically, adenocarcinoma, mesothelioma,  
 CC ovarian carcinoma, pancreatic carcinoma or non-small cell lung carcinoma.  
 CC  
 CC Sequence 399 AA:  
 SQ  
 Query Match 46.9%; Score 1531; DB 21; Length 399;  
 Best Local Similarity 97.1%; Pred. No. 2.2e-131;  
 Matches 298; Conservative 1; Mismatches 0; Indels 8; Gaps 1;  
 OY 294 EYEKTRACPSGKKAREIDSLIFKKWELEACVDALATQMDRVNAIPFTYEQDLYLKH 353  
 DB 1 evektacpsgkkaareidesliffkkweleacvdaallatqmdrvnaipfyeqldv1kh 60  
 OY 354 LDELYPGYPSVSIQHLGYLFLKMSPEDIRKWNVTSLETALKALLEVDKGHEMSPOAPRR 413  
 DB 61 ldelypgypesvsiqhlgyilfkmspedirkwnvtsletalkallevnkghems----- 113  
 OY 414 LPOVATLIDRFVKGKGGQLDKDLDLTLTAFYPGYLCSSPEELSSVPSSIAVVRPQD 473  
 DB 114 -pqvatllidrfvkggrgldkldtltltafypgylcsspeelssvpsaiwvrpqd 172  
 OY 474 CDRPOLDVLYPKARLAFQNMNGSEYFVKIQSFLGAPREDIKALSOQNVSMDLATFM 533  
 DB 173 cdrpqldvlypkarlaifqnmngseyfvkigsflggapredikalsgqnvsmldatfm 232  
 OY 534 TDVALPLTVAEVQKLLGPHVEGLKAEERHRRPVDMILRORODDLTLGLGOGGIPN 593  
 DB 233 tdavplltvaevqkllgphveglkaeerhrpvrdmllrqrddldtlglg9gipngyl 292  
 OY 594 VLDLSVQ 600  
 DB 293 vldlsvq 299  
 RESULT 6  
 ID AAR53991 standard; peptide: 65 AA.  
 AC AAR53991:  
 XX  
 DT 08-DEC-1994 (first entry)  
 XX  
 DE Meg-Pot fragment.  
 XX  
 KW Megakaryocyte potentiator; Meg-Pot; thrombocytopoiesis;  
 KW platelet; amplification; primer; polymerase chain reaction; PCR.  
 XX  
 OS Homo sapiens.

XX MO9410312-A.  
XX 11-MAY-1994.  
XX  
XX  
XX 25-OCT-1993; 93WO-JP01540.  
XX  
XX 23-OCT-1992; 92JP-0286153.  
XX 11-NOV-1992; 92JP-0301387.  
XX 09-DEC-1992; 92JP-0329546.  
XX  
XX (CHUS ) CHUGAI SEIYAKU KK.  
XX  
XX Hattori K, Kojima T, Oh-eda M, Yamaguchi N;  
XX WPI: 1994-167467/20.  
XX N-PSDB: AA063971.  
XX  
XX New mega:karyocyte potentiator - for potential treatment of  
XX thrombocytopenia  
XX  
XX Disclosure: Page 51; 74pp; Japanese.  
XX  
XX Use of the primers given in AA077809-10 in PCR resulted in the  
XX fragments given in AA063971.  
XX DNA encoding Meg-Pot has potential use in treatment of  
XX thrombocytopenia and low platelet function.  
XX  
XX Sequence 65 AA:

Query Match 10.0%; Score 325; DB 15; Length 65;  
Best Local Similarity 100.0%; Pred. No. 2.8e-22;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 LIGFPCAEEVSGISTEERVELAVALAOKNVKLTSEQLRCIAHRLSEPPEDLDALPDLULLF 123  
DB 1 lllgfcaevsgistervrelavalaqknvklstegrlclahrlseppedldalpidlllf 60

QY 124 LNPDA 128  
DB 61 lnpda 65

## RESULT 7

AA053988  
ID AA053988 standard; peptide: 35 AA.

XX  
AC AA053988;

XX  
DT 08-DEC-1994 (first entry)

XX  
DE Meg-Pot N-terminal peptide.

XX  
KW Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;

XX  
KW Platelet.

XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers

FT MISC-difference 19 /note= "residue not identified"

XX  
PN MO9410312-A.

XX  
PD 11-MAY-1994.

XX  
PF 25-OCT-1993; 93WO-JP01540.

XX  
PR 23-OCT-1992; 92JP-0286153.

XX  
PR 11-NOV-1992; 92JP-0301387.

XX  
PR 09-DEC-1992; 92JP-0329546.

PA (CHUS ) CHUGAI SEIYAKU KK.  
XX  
XX Hattori K, Kojima T, Oh-eda M, Yamaguchi N;  
XX WPI: 1994-167467/20.  
XX  
XX New mega:karyocyte potentiator - for potential treatment of  
XX thrombocytopenia  
XX  
XX Disclosure: Page 49; 74pp; Japanese.  
XX  
XX DNA encoding Meg-Pot has potential use in treatment of  
XX thrombocytopenia and low platelet function.  
XX  
XX Sequence 35 AA:

Query Match 5.0%; Score 163; DB 15; Length 35;  
Best Local Similarity 97.1%; Pred. No. 7.1e-08;  
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 83 LAVALAOKNVKISTEQLRCIAHRLSEPPEDLDALP 117  
DB 1 lavalagknvklstegrlclahrlseppedldalp 35

## RESULT 8

AA053990  
ID AA053990 standard; peptide: 25 AA.

XX  
AC AA053990;

XX  
DT 08-DEC-1994 (first entry)

XX  
DE Meg-Pot fragment.

XX  
KW Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;

XX  
KW Platelet; amplification; primer: polymerase chain reaction; PCR.

XX  
OS Homo sapiens.

XX  
PN MO9410312-A.

XX  
PD 11-MAY-1994.

XX  
PF 25-OCT-1993; 93WO-JP01540.

XX  
PR 23-OCT-1992; 92JP-0286153.

XX  
PR 11-NOV-1992; 92JP-0301387.

XX  
PR 09-DEC-1992; 92JP-0329546.

XX  
PA (CHUS ) CHUGAI SEIYAKU KK.

XX  
PI Hattori K, Kojima T, Oh-eda M, Yamaguchi N;

XX  
DR WPI: 1994-167467/20.

XX  
DR N-PSDB: AA063970.

XX  
PT New mega:karyocyte potentiator - for potential treatment of

XX  
PT thrombocytopenia

XX  
PS Disclosure: Page 50; 74pp; Japanese.

XX  
CC Use of the primers given in AA077805-08 in PCR resulted in the

XX  
CC fragments given in AA063969-70.

XX  
CC DNA encoding Meg-Pot has potential use in treatment of

XX  
CC thrombocytopenia and low platelet function.

XX  
SQ Sequence 25 AA:

Query Match 3.9%; Score 128; DB 15; Length 25;  
Best Local Similarity 100.0%; Pred. No. 6.7e-05;



Matches	25; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
OY	106	LSEPPEDALPLDLLFPNPAFS	130		
Db	1	lseppeidalpldllfnpafsf	25		
RESULT	9				
AAE03600					
ID	AAE03600	standard: Protein; 794 AA.			
XX	AAE03600;				
AC					
XX					
DT	07-AUG-2001	(first entry)			
XX					
DE	Human	leucine-rich repeat-containing protein, AZAD.			
XX					
KW	Human;	leucine-rich repeat; AZAD; neurodegenerative disorder;			
KW	CNS disorder;	central nervous system disorder; prostate disorder;			
KW	prostatitis;	benign prostatic hyperplasia; adenocarcinoma;			
KW	prostate cancer;	genitourinary system carcinoma; testicular tumour;			
KW	Alzheimer's disease;	dementia; Parkinson's disease; multiple sclerosis;			
KW	amyotrophic lateral sclerosis;	epilepsy; psychiatric disorder; mania;			
KW	depression;	schizophrenia; anxiety; phobic disorder; learning disorder;			
KW	memory disorder;	amnesia; migraine; protein-protein interaction;			
KW	cellular activity;	neoplastic transformation; drug screening;			
KW	forensic identification;	gene therapy.			
XX					
OS	Homo sapiens.				
XX					
PH	Key	Location/Qualifiers			
FT	Peptide	1..58			
FT		/label= Signal_peptide			
FT	Protein	59..794			
FT		/label= Mature_AZAD_protein			
FT	Region	65..94			
FT		/note= "N-terminal leucine-rich repeat (LRR)"			
FT	Region	427..456			
FT		/note= "N-terminal leucine-rich repeat (LRR)"			
FT	Region	96..119			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	120..143			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	144..167			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	168..191			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	192..215			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	216..239			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	240..263			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	264..287			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	288..311			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	312..333			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	458..481			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	482..505			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	506..529			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	530..553			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	554..577			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	578..601			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	602..625			
FT		/note= "Leucine-rich repeat (LRR)"			

FT Region 626..649  
 FT /note= "Leucine-rich repeat (LRR)"  
 FT Region 651..674  
 FT /note= "Leucine-rich repeat (LRR)"  
 FT Region 676..697  
 FT /note= "Leucine-rich repeat (LRR)"  
 FT Region 698..719  
 FT /note= "Leucine-rich repeat (LRR)"  
 FT Region 707..755  
 FT /note= "C-terminal leucine-rich repeat (LRR)"  
 XX  
 XX WO200142286-A2.  
 XX  
 XX 14-JUN-2001.  
 XX  
 XX 07-DEC-2000; 2000WO-US33140.  
 XX  
 XX 08-DEC-1999; 99US-0456592.  
 XX  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX  
 XX Khodadoust MM;  
 XX  
 XX WPI: 2001-381633/40.  
 XX  
 XX N-PSDB: AAD08044.  
 DR  
 DR  
 PT New leucine-rich repeat-containing protein and nucleic acid molecules  
 PT for diagnosing, treating neural disorders, such as neurodegenerative  
 PT disorders, such as Alzheimer's disease, dementia, epilepsy and prostate  
 PT cancer -  
 XX  
 XX  
 XX  
 XX  
 PS  
 PS  
 XX  
 XX  
 CC The present sequence is human leucine-rich repeat-containing  
 CC secreted protein, AZAD. AZAD polypeptide is useful for  
 CC identifying a compound which modulates its activity and binds to it.  
 CC AZAD proteins and nucleic acid molecules are useful for treating and  
 CC diagnosing AZAD-mediated or related disorders, which includes a neural  
 CC disorder (e.g. neurodegenerative disorders, including CNS disorders)  
 CC and a prostate disorder, e.g. prostatitis, benign prostatic hyperplasia  
 CC and cancer (e.g. adenocarcinoma, prostate cancer, genitourinary system  
 CC carcinomas and testicular tumour). The neurodegenerative disorders  
 CC include Alzheimer's disease, dementias related to Alzheimer's disease,  
 CC Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis,  
 CC epilepsy, psychiatric disorders e.g. depression, schizophrenia  
 CC disorders, mania, anxiety or phobic disorders, learning or memory  
 CC disorders e.g. amnesia or age-related memory loss, and neurological  
 CC disorders e.g. migraine. AZAD molecules are useful as markers of  
 CC for predisposition of disease states, as markers for precursors of disease state,  
 CC pharmacogenomic profile of a subject. The AZAD polypeptide and  
 CC polynucleotide are capable of modulating protein-protein interaction,  
 CC e.g. by interacting with an extracellular component, thereby modulating  
 CC cellular activities, including attachment, adhesion, migration,  
 CC patterning, growth and/or differentiation of a cell. AZAD proteins  
 CC regulate embryonic development and differentiation, tissue maintenance  
 CC and function, pathological conditions, e.g. neuronal degeneration,  
 CC neoplastic transformation and tumour progression. AZAD proteins are  
 CC useful as immunogens to raise anti-AZAD antibodies which are useful to  
 CC detect and isolate AZAD proteins and modulate its activity. AZAD  
 CC proteins are useful to screen for naturally occurring AZAD substrates  
 CC and to screen for drugs or compounds which modulate AZAD activity.  
 CC AZAD nucleic acid fragments are useful as primers or hybridisation  
 CC probes for the detection of AZAD-encoding nucleic acids. cDNA encoding  
 CC the AZAD protein are useful in gene therapy. AZAD sequences are also  
 CC useful to map their respective genes on a chromosome, for tissue typing  
 CC and in forensic identification of a biological sample.  
 CC  
 XX  
 XX  
 SQ Sequence 794 AA;

Query Match 3.8%; Score 123.5; DB 22; Length 794;  
 Best Local Similarity 29.0%; Pred. No. 0.037;





```
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144337.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.58; Score 114.5; DB 21; Length 349;
Best Local Similarity 20.78; Pred. No. 0.069;
Matches 92; Conservative 63; Mismatches 128; Indels 161; Gaps 23;

QY 79 RVRELAVALAQNKLSTEQRLCLA--HRISEPPEDDALPLDLLFLNPAFSGPOACT 136
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 rvrm--klsgqglevaeqgcmalserygapkpetadial-----lhainsg----v 54

QY 137 RFESEIRKANVDLPRCAPERORLLPAAACGVR-----GSLISEADVVALGCLA 187
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 tff-----dstadmygpecnelllgkalkdykvelatkigffivegeisevrg-- 104

QY 188 CDLPGRFVAESAENVLLPL--VSCPG-----PLDQDQGEAARALOGG----GPP 231
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 105 ---dpeyvraeaslkrlidiacidllyqhndtrvpieitumrelkklveegkikygl 161

QY 232 YGPST-----WSTMDALRGLLPV---LGOPFIIRSP--OGIVAAMR 270
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 162 easastlirhavnpitavqlewsrdaeedlhploreigjlvayspjrgflaagp 221

QY 271 Q--RSSRDPKWRQPERITLRRFRREVEKTAQPSGKAREIDESLIRYKK---WELBAC 324
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 klenlendfirkattl--prfge-----nvdhnhkklfeksamaekgyc 266
```

PR	04-JUN-1999	9905-0137752
PR	07-JUN-1999	9905-0137720
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139452
PR	16-JUN-1999	9905-0139453
PR	17-JUN-1999	9905-0139492
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139750
PR	18-JUN-1999	9905-0139763
PR	18-JUN-1999	9905-0139817
PR	21-JUN-1999	9905-0139817
PR	22-JUN-1999	9905-0139899
PR	23-JUN-1999	9905-0140353
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140695
PR	28-JUN-1999	9905-0140823
PR	29-JUN-1999	9905-0140991
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141842
PR	01-JUL-1999	9905-0142154
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142390
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0142820
PR	12-JUL-1999	9905-0142877
PR	13-JUL-1999	9905-0143542
PR	14-JUL-1999	9905-0143625
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144333
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144335
PR	20-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144353
PR	20-JUL-1999	9905-0144484
PR	21-JUL-1999	9905-0144814
PR	21-JUL-1999	9905-0145086
PR	21-JUL-1999	9905-0145088
PR	22-JUL-1999	9905-0145085
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145088
PR	22-JUL-1999	9905-0145142
PR	23-JUL-1999	9905-0145145
PR	23-JUL-1999	9905-0145218
PR	23-JUL-1999	9905-0145224
PR	26-JUL-1999	9905-0145576
PR	27-JUL-1999	9905-0145513
PR	27-JUL-1999	9905-0145518
PR	27-JUL-1999	9905-0145519
PR	28-JUL-1999	9905-0145561
PR	02-AUG-1999	9905-0145686
PR	02-AUG-1999	9905-0146588
PR	02-AUG-1999	9905-0146589
PR	03-AUG-1999	9905-0147038
PR	04-AUG-1999	9905-0147204
PR	04-AUG-1999	9905-0147302
PR	05-AUG-1999	9905-0147260
PR	05-AUG-1999	9905-0147260



CC fragments given in AA063969-70.  
CC DNA encoding Meg-Pot has potential use in treatment of  
CC thrombocytopenia and low platelet function.

XX Sequence 23 AA:

Query Match 3.5%; Score 114; DB 15; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.0011;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 62 ROLGFPCEAVSGSLSTERVRELA 84  
Db 1 RGLGFPCEAVSGSLSTERVRELA 23

## RESULT 15

AA051557  
ID AA051557 standard; Protein; 913 AA.

XX AA051557;

XX 18-MAY-2000 (first entry)

DE Human PLA2 protein.

XX PLA2; phospholipase A2; phosphatide 2-acyl hydrolase; human; therapy;  
KW arachidonic acid; lysophospholipid; Alzheimer's disease.

XX Homo sapiens.

XX US6025178-A.

PN 15-FEB-2000.

PD 28-MAR-1997; 97US-0827208.

PE 29-MAR-1996; 96US-0014608.

PR (ELIL ) LILLY & CO ELI.

PI Sharp JD, Scriffler BA, Cholu XC, Kramer RM, Pickard RT;

DR WPI; 2000-181816/16.

DR N-PSDB; AA288756; AA288757.

PT An isolated amino acid having phospholipase (PL)A2 activity is useful  
PT in assays to identify inhibitors having a therapeutic benefit, such as  
PT inhibiting the central role of PLA2 in the inflammatory component of  
PT Alzheimer's disease -

PS Claim 1; Column 53-58; 32pp; English.

XX This invention describes a novel human phospholipase A2 (PLA2) protein  
XX (1) and its encoding nucleic acid. The amino acid (1) releases  
XX arachidonic acid in specific tissues characterized by unique membrane  
XX phospholipids, by generating lysophospholipid species which are  
XX deleterious to membrane integrity or by remodeling of unsaturated species  
XX of membrane phospholipids through deacylation/reacylation mechanisms. The  
XX amino acid is useful in assays to identify inhibitors having a  
XX therapeutic benefit, such as inhibiting the central role of PLA2 in the  
XX inflammatory component of Alzheimer's disease. The amino acid (1) allows  
XX sensitive and rapid screening and identification of inhibitors of  
XX phospholipase A2. This sequence represents the human PLA2 protein (also  
XX known as phosphatide 2-acyl hydrolase).

XX Sequence 913 AA;

Query Match 3.5%; Score 113.5; DB 21; Length 913;  
Best Local Similarity 20.8%; Pred. No. 0.38;  
Matches 143; Conservative 72; Mismatches 257; Indels 215; Gaps 33;

OY 1 MALORLDEPCWSCGDRPSLLFLFLFSLGAVHWHPARTLAGETGES---APLG-GVLTTPPHNI 56  
Db melkfvddqdlvvgddp---vlsvlfaqltr-----agefirresflspgqgrilevelrfl 260  
OY 57 SLSLPRÖLLGFPCEAVSGSLSTERVRELAVALAÖKNVKLSTEDÖRLCLARLSBPPEDDAL 116  
Db 261 gsladr-----gewlvs-----ngvlvarelscldhvgleedgqkse 298  
OY 117 PLDLFLFNPDAFSGPQACT-----RFE-SRTTANVDLLPRGAPREQRLLPALAÖWG 169  
Db 299 hrvgllv---pasccepgqasvgtgtrfthpcacwqeqlstlrlqadeeqkxaplsalpsg 356  
OY 170 -----VRGSLSEADVR---AUGSLACDLPGRFVESAVAL---LPRLVS 208  
Db 357 gvvrlvftsgqplmrvelkkeaqlrelavrlgfgpcaseeqaflstrkqvvaalrlqalq 416  
OY 209 CPGLPDÖDQÖEAKRALÖGGÖPÖPGPÖSTWSVSTÖDALRGLLPVLGÖ---PIRSIPÖG 264  
Db 417 lddldqedeiprvaimatggy-----lramtslygqlaglkqlldcvsyltga 466  
OY 265 IVAAWRÖRS-SHÖPSMRÖPERTILRFRREVEKÖRACPSGKKAREIDSLIFÖKKWÖLEA 323  
Db 467 sgstwalanlyedpewsqkd-----laqpt----- 491  
OY 324 CVDAALLATÖMDRVNVALPÖTÖEDVLÖKÖLDELYPÖGYPÖSVIÖHLGÖYFLÖKMSPEDIR 383  
Db 492 ---eliktgyknlgylappsqldgrylgelaerarlgypsclt----- 531  
OY 384 KWNVTSLETIKALLEVDGHEMSPOAF-----RRLPÖVATLIDRFVKGÖGÖLDÖKÖTLDÖT 438  
Db 532 -nlwallneallndepdhklsdqrealshgqndlpilycal-----ntkqg-----s 577  
OY 439 LTAÖPÖGYLÖSTSPÖELÖSSVPPÖSSIMAVRÖDÖLÖDÖCD---PROÖDVLÖYKARLAF-ÖNÖNG 495  
Db 578 lttfegwecsfpyevf---pkysgafipöselöföfömgölmökröpestöicöföegöts 634  
OY 496 SEYFVKIÖSFL--GGAPÖE--DLKALSOÖQVNSMDIATÖMKLR-----TDAVL 538  
Db 635 nlyaanldqsllywaspepöqfdwrvrnqandkeqvlkkieepstagrölaeffötdöilt 694  
OY 539 --PLTVAEVÖKÖLGPÖHVEGLKAEERÖRÖPÖVÖRWÖILÖRÖDÖDÖLÖTÖGLÖGÖGÖIPN----- 590  
Db 695 wrplagathnflröghfn--kdyföghpöfstv---kattld-----ölpnqölpöse 740  
OY 591 -----GYVLÖDLSVÖETLSGTPÖCL 609  
Db 741 phclldövgyll-----ntöcl 757

Search completed: December 7, 2001, 10:01:37  
Job time: 720 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: December 7, 2001, 09:48:52 ; Search time 22.61 Seconds  
(Without alignments)  
625.036 Million cell updates/sec

Title: US-09-215-035-2  
Perfect score: 3261  
Sequence: 1 MALQRHDPWCGDRPGSL.....LLGPGVLTVALLASTIA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3261	100.0	628	US-08-776-271-2	Sequence 2, Appl1
2	3261	100.0	628	US-09-215-035-2	Sequence 2, Appl1
3	3055	93.7	622	US-08-426-819A-35	Sequence 35, Appl1
4	2956	90.6	584	US-08-426-819A-36	Sequence 36, Appl1
5	1247	38.2	248	US-08-426-819A-37	Sequence 37, Appl1
6	325	10.0	65	US-08-426-819A-33	Sequence 33, Appl1
7	210	6.4	40	US-08-426-819A-19	Sequence 19, Appl1
8	163	5.0	35	US-08-426-819A-5	Sequence 5, Appl1
9	134	4.1	6095	US-09-144-085-2	Sequence 2, Appl1
10	113.5	3.5	778	US-09-460-145-2	Sequence 2, Appl1
11	113.5	3.5	797	US-08-827-208-3	Sequence 3, Appl1
12	113.5	3.5	913	US-09-500-358-3	Sequence 3, Appl1
13	113.5	3.5	913	US-09-498-809-3	Sequence 3, Appl1
14	113.5	3.5	913	US-08-426-819A-18	Sequence 18, Appl1
15	111	3.4	33	US-08-762-428A-6	Sequence 6, Appl1
16	107.5	3.3	1288	US-07-642-734C-4	Sequence 4, Appl1
17	107.5	3.3	3567	US-08-439-009A-4	Sequence 4, Appl1
18	107.5	3.3	3567	US-08-339-152A-17	Sequence 17, Appl1
19	106.5	3.3	634	US-08-339-152A-16	Sequence 16, Appl1
20	106.5	3.3	653	US-08-007-999B-3	Sequence 3, Appl1
21	106.5	3.3	653	US-08-689-276A-3	Sequence 3, Appl1
22	105.5	3.3	653	US-08-026-138E-3	Sequence 2, Appl1
23	105	3.2	1239	PCT-US95-14418-2	Sequence 2, Appl1
24	103.5	3.2	834	PCT-US95-15327-2	Sequence 2, Appl1
25	103.5	3.2	834	US-07-977-434-10	Sequence 10, Appl1
26	101.5	3.1	834	US-08-073-384C-6	Sequence 6, Appl1
27	101.5	3.1	834		

28	101.5	3.1	834	1	US-08-254-359A-6	Sequence 6, Appl1
29	101.5	3.1	834	1	US-08-384-490-31	Sequence 31, Appl1
30	101.5	3.1	834	1	US-08-483-043-6	Sequence 6, Appl1
31	101.5	3.1	834	1	US-08-459-383-31	Sequence 31, Appl1
32	101.5	3.1	834	1	US-08-458-819-10	Sequence 10, Appl1
33	101.5	3.1	834	1	US-08-481-238-6	Sequence 6, Appl1
34	101.5	3.1	834	2	US-08-471-066B-6	Sequence 6, Appl1
35	101.5	3.1	834	2	US-08-484-956-6	Sequence 6, Appl1
36	101.5	3.1	834	2	US-08-757-653-6	Sequence 6, Appl1
37	101.5	3.1	834	2	US-08-589-491-6	Sequence 6, Appl1
38	101.5	3.1	834	2	US-08-756-386-6	Sequence 6, Appl1
39	101.5	3.1	834	2	US-08-823-516-6	Sequence 6, Appl1
40	101.5	3.1	834	3	US-08-682-853A-6	Sequence 6, Appl1
41	101.5	3.1	834	3	US-08-759-038-6	Sequence 6, Appl1
42	101.5	3.1	834	3	US-08-758-314-6	Sequence 6, Appl1
43	101.5	3.1	834	5	PCT-US91-07035-10	Sequence 10, Appl1
44	100.5	3.1	609	4	US-08-980-115-11	Sequence 11, Appl1
45	99.5	3.1	514	4	US-09-385-028-5	Sequence 5, Appl1

## ALIGNMENTS

RESULT 1  
: Sequence 2, Application US/08776271  
: Patent No. 6083502  
: GENERAL INFORMATION:  
: APPLICANT: Pastan, Ira  
: TITLE OF INVENTION: Mesothelin, a Differentiation Antigen  
: TITLE OF INVENTION: Present on Mesothelioma, Mesotheliomas and Ovarian Cancers  
: TITLE OF INVENTION: and Methods and Kits for Targeting the Antigen  
: NUMBER OF SEQUENCES: 8  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Townsend and Townsend and Crew LLP  
: STREET: Two Embarcadero Center, Eighth Floor  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94111-3834  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/776,271  
: FILING DATE: 01-DEC-1998  
: CLASSIFICATION: 424  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: WO PCT/US97/00224  
: FILING DATE: 03-JAN-1997  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 60/010,166  
: FILING DATE: 05-JAN-1996  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Paris, Susan K. 41,739  
: REGISTRATION NUMBER: 015280-25910005  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (415) 576-0200  
: TELEFAX: (415) 576-0300  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 628 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-776-271-2  
Query Match 100.0%; Score 3261; DB 3; Length 628;



Best Local Similarity 100.0%; Pred. No. 1.5e-303;  
Matches 628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALORLDCWSCGDRGSLFLFLSLGWHPARTLAGETGESAPLGGVLTTPHNISLS 60  
DB 1 MALORLDCWSCGDRGSLFLFLSLGWHPARTLAGETGESAPLGGVLTTPHNISLS 60  
QY 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNKVLTSTEDLRCIAHRLSEPPEDLALPDL 120  
DB 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNKVLTSTEDLRCIAHRLSEPPEDLALPDL 120  
QY 121 LFLNPDASFSGQACTRFSRTTKANVDLLPRGAPERORLLPALACWVSGSLSEADY 180  
DB 121 LFLNPDASFSGQACTRFSRTTKANVDLLPRGAPERORLLPALACWVSGSLSEADY 180  
QY 181 RALGGLACDLPRFVAESAENVLLPRLVSCPGPLDDQDEAARAALOGGPPYGPSTWSV 240  
DB 181 RALGGLACDLPRFVAESAENVLLPRLVSCPGPLDDQDEAARAALOGGPPYGPSTWSV 240  
QY 241 STMALRGILLPVLGOPITRSIPQGIWAAMRORSSRDPSPROBERTILRRFRREVEKTAC 300  
DB 241 STMALRGILLPVLGOPITRSIPQGIWAAMRORSSRDPSPROBERTILRRFRREVEKTAC 300  
QY 301 PSGKAREIDESLIFYKKWELEACVDALLATQMDRVNAIPFTYEQDLVLAHKHDELTPQ 360  
DB 301 PSGKAREIDESLIFYKKWELEACVDALLATQMDRVNAIPFTYEQDLVLAHKHDELTPQ 360  
QY 361 GYPSVIOHGLFLFKMSPEDIRKNNVTSLETALKALEVDKHEMSPOAPRRRLPOVATL 420  
DB 361 GYPSVIOHGLFLFKMSPEDIRKNNVTSLETALKALEVDKHEMSPOAPRRRLPOVATL 420  
QY 421 IDRFYKRGQLDKDTLDTLTAFAFYGLCSLSPBELSVSPSSINAVRPQDDTCDPRQD 480  
DB 421 IDRFYKRGQLDKDTLDTLTAFAFYGLCSLSPBELSVSPSSINAVRPQDDTCDPRQD 480  
QY 481 VLYPKARLAFQNMNGSEFVATIOSFLGAPTEBDLKALSOQVNSMDLATFMKRTDAVPL 540  
DB 481 VLYPKARLAFQNMNGSEFVATIOSFLGAPTEBDLKALSOQVNSMDLATFMKRTDAVPL 540  
QY 541 TVAEYQKLLGPHVEGLKAEEHRHPYRDWILRQRODDLTGLGOGGIPNGYLVLDLSVQ 600  
DB 541 TVAEYQKLLGPHVEGLKAEEHRHPYRDWILRQRODDLTGLGOGGIPNGYLVLDLSVQ 600  
QY 601 ETLSGTPCLLGGPVLYVALLLASTLA 628  
DB 601 ETLSGTPCLLGGPVLYVALLLASTLA 628

RESULT 2  
US-09-215-035-2  
Sequence 2, Application US/09215035  
Patent No. 6153430  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
TITLE OF INVENTION: Mesothelin, a Differentiation Antigen  
TITLE OF INVENTION: Present on Mesothelioma, Mesotheliomas and Ovarian Cancers  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/215,035

FILING DATE: No. 6153430 yet assigned  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/776,271  
FILING DATE: 01-DEC-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/00224  
FILING DATE: 03-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/010,166  
FILING DATE: 05-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Faris, Susan K.  
REGISTRATION NUMBER: 41,739  
REFERENCE/DOCKET NUMBER: 015280-259110US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 628 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-215-035-2

Query Match 100.0%; Score 3261; DB 4; Length 628;  
Best Local Similarity 100.0%; Pred. No. 1.5e-303;  
Matches 628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALORLDCWSCGDRGSLFLFLSLGWHPARTLAGETGESAPLGGVLTTPHNISLS 60  
DB 1 MALORLDCWSCGDRGSLFLFLSLGWHPARTLAGETGESAPLGGVLTTPHNISLS 60  
QY 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNKVLTSTEDLRCIAHRLSEPPEDLALPDL 120  
DB 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNKVLTSTEDLRCIAHRLSEPPEDLALPDL 120  
QY 121 LFLNPDASFSGQACTRFSRTTKANVDLLPRGAPERORLLPALACWVSGSLSEADY 180  
DB 121 LFLNPDASFSGQACTRFSRTTKANVDLLPRGAPERORLLPALACWVSGSLSEADY 180  
QY 181 RALGGLACDLPRFVAESAENVLLPRLVSCPGPLDDQDEAARAALOGGPPYGPSTWSV 240  
DB 181 RALGGLACDLPRFVAESAENVLLPRLVSCPGPLDDQDEAARAALOGGPPYGPSTWSV 240  
QY 241 STMALRGILLPVLGOPITRSIPQGIWAAMRORSSRDPSPROBERTILRRFRREVEKTAC 300  
DB 241 STMALRGILLPVLGOPITRSIPQGIWAAMRORSSRDPSPROBERTILRRFRREVEKTAC 300  
QY 301 PSGKAREIDESLIFYKKWELEACVDALLATQMDRVNAIPFTYEQDLVLAHKHDELTPQ 360  
DB 301 PSGKAREIDESLIFYKKWELEACVDALLATQMDRVNAIPFTYEQDLVLAHKHDELTPQ 360  
QY 361 GYPSVIOHGLFLFKMSPEDIRKNNVTSLETALKALEVDKHEMSPOAPRRRLPOVATL 420  
DB 361 GYPSVIOHGLFLFKMSPEDIRKNNVTSLETALKALEVDKHEMSPOAPRRRLPOVATL 420  
QY 421 IDRFYKRGQLDKDTLDTLTAFAFYGLCSLSPBELSVSPSSINAVRPQDDTCDPRQD 480  
DB 421 IDRFYKRGQLDKDTLDTLTAFAFYGLCSLSPBELSVSPSSINAVRPQDDTCDPRQD 480  
QY 481 VLYPKARLAFQNMNGSEFVATIOSFLGAPTEBDLKALSOQVNSMDLATFMKRTDAVPL 540  
DB 481 VLYPKARLAFQNMNGSEFVATIOSFLGAPTEBDLKALSOQVNSMDLATFMKRTDAVPL 540  
QY 541 TVAEYQKLLGPHVEGLKAEEHRHPYRDWILRQRODDLTGLGOGGIPNGYLVLDLSVQ 600  
DB 541 TVAEYQKLLGPHVEGLKAEEHRHPYRDWILRQRODDLTGLGOGGIPNGYLVLDLSVQ 600  
QY 601 ETLSGTPCLLGGPVLYVALLLASTLA 628  
DB 601 ETLSGTPCLLGGPVLYVALLLASTLA 628

Db 601 ETLSTGTPCLLGPVLTVALALLASTLA 628

RESULT 3

US-08-426-819A-35

Sequence 35, Application US/08426819A

Patent No. 5723318

GENERAL INFORMATION:

APPLICANT: Yamaguchi, No. 5723318oml

APPLICANT: Kojima, Tetsuo

APPLICANT: Oh-Eda, Masayoshi

APPLICANT: Hatcort, Kunihiro

TITLE OF INVENTION: Genes Coding for Megakaryocyte

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/426, 819A

FILING DATE: 21-Apr-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 230-107P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-205-8000

TELEFAX: 703-205-8050

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:

LENGTH: 622 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-426-819A-35

Query Match 93.78; Score 3055; DB 1; Length 622;

Best Local Similarity 93.48; Pred. No. 7.8e-284;

Matches 601; Conservative 3; Mismatches 16; Indels 10; Gaps 3;

QY 1 MALRLDP-CWSCGDRP-GSLILFLFSLGVHPARTLAGETGTSAPLGVLTTPPHN1SS 58

Db 1 MALTPARILGSCGTFRALGSLIFLFLFSLGVHPARTLAGETGTSAPLGVLTTPPHN1SS 60

QY 59 LSPQQLGFPCEAVESGLSTERYRELAVALAOKNKLSTEQRLCLAHRLSEPPEDDALPL 118

Db 61 LSPQQLGFPCEAVESGLSTERYRELAVALAOKNKLSTEQRLCLAHRLSEPPEDDALPL 120

QY 119 DLLFLNPDASSGQACRFRTSRITKANVDLLPRGAPRQRLPAALACNVRGSLSEA 178

Db 121 DLLFLNPDASSGQACRFRTSRITKANVDLLPRGAPRQRLPAALACNVRGSLSEA 180

QY 179 DYRALGGLACDLPGRFVASEVLLPRLVSCPGPLDDQOQANARALGGGPPGPPSTW 238

Db 181 DYRALGGLACDLPGRFVASEVLLPRLVSCPGPLDDQOQANARALGGGPPGPPSTW 240

QY 239 SVSTMDALRGLLPVAGOPRIIRSIPQGIYAANRORSSRDPMSROBERTILRPRFRREVEKT 298

Db 241 SVSTMDALRGLLPVAGOPRIIRSIPQGIYAANRORSSRDPMSROBERTILRPRFRREVEKT 300

QY 299 ACPGSKARREIDESIIFPKKWEACVDAALLATQMDRVNAIPTTYEQLDVKKHLDLY 358

Db 301 ACPGSKARREIDESIIFPKKWEACVDAALLATQMDRVNAIPTTYEQLDVKKHLDLY 360

QY 359 POGYPSYIOHLYFLTKMSPEDIRKMNVTSLFTKALLEVDKHEMSPOARRPLPOVA 418

Db 361 POGYPSYIOHLYFLTKMSPEDIRKMNVTSLFTKALLEVDKHEMSPOARRPLPOVA 412

QY 419 TLIDRFVGRQQLDKDTLDTLTAFFPGYICLSPEELSSVPSSIMAVRPDDLTCDPRQ 478

Db 413 TLIDRFVGRQQLDKDTLDTLTAFFPGYICLSPEELSSVPSSIMAVRPDDLTCDPRQ 472

QY 479 LDVLYPKARLAFQNNNGSEYVKIQSEFLGAPTEDLKALSOONVSMDLATPMKLRTPDAVL 538

Db 473 LDVLYPKARLAFQNNNGSEYVKIQSEFLGAPTEDLKALSOONVSMDLATPMKLRTPDAVL 532

QY 539 PLTYAEVOKLGPVHEGKAEERHPRVMDWTLRQRODDLTGLGGLGNGYLVLDLS 598

Db 533 PLTYAEVOKLGPVHEGKAEERHPRVMDWTLRQRODDLTGLGGLGNGYLVLDLS 592

QY 599 VOETLSTGTPCLLGPVLTVALALLASTLA 628

Db 593 VOETLSTGTPCLLGPVLTVALALLASTLA 622

RESULT 4

US-08-426-819A-36

Sequence 36, Application US/08426819A

Patent No. 5723318

GENERAL INFORMATION:

APPLICANT: Yamaguchi, No. 5723318oml

APPLICANT: Kojima, Tetsuo

APPLICANT: Oh-Eda, Masayoshi

APPLICANT: Hatcort, Kunihiro

TITLE OF INVENTION: Genes Coding for Megakaryocyte

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/426, 819A

FILING DATE: 21-Apr-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 230-107P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-205-8000

TELEFAX: 703-205-8050

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 584 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: protein

HYPOTHETICAL: NO

FRAGMENT TYPE: internal

US-08-426-819A-36

Query Match 90.68; Score 2956; DB 1; Length 584;

Best Local Similarity 97.38; Pred. No. 2.1e-274;

Matches 576; Conservative 2; Mismatches 6; Indels 8; Gaps 1;

QY 37 GETGESAPLGVLTTPHNIISSLSPROLLGFPACAEVSGLSTERYRELAVALAOKNWKST 96  
1 GETGEAPAPLDGVLANPNNISSLSPROLLGFPACAEVSGLSTERYRELAVALAOKNWKST 60  
QY 97 EQLRCLARHLSPEPDDLPDLPLDLFLNPDASFSGPQACTRFFSRITKANVDLPRGAP 156  
61 EQLRCLARHLSPEPDDLPDLPLDLFLNPDASFSGPQACTRFFSRITKANVDLPRGAP 120  
QY 157 RORLLPALACMGVAGSLSEADVAGLACDLPGRVAESAELVLRVSCPEPLD 216  
121 RORLLPALACMGVAGSLSEADVAGLACDLPGRVAESAELVLRVSCPEPLD 180  
QY 217 OOEAAARALOGGPPYGPSTWSVTMDALRGLLPVLGOPITRSIPQIVAMRORSSRD 276  
181 OOEAAARALOGGPPYGPSTWSVTMDALRGLLPVLGOPITRSIPQIVAMRORSSRD 240  
QY 277 PSWROPERTILRPFRREVETKACPSGKKAREIDSLIFPKKWELEACVDALLATQMDR 336  
241 PSWROPERTILRPFRREVETKACPSGKKAREIDSLIFPKKWELEACVDALLATQMDR 300  
QY 337 VNAPFTYEQDLVKHKKDELTPQGYPSVYQHLGYLFLKMSPEDIRKWNVTSLTLAL 396  
301 VNAPFTYEQDLVKHKKDELTPQGYPSVYQHLGYLFLKMSPEDIRKWNVTSLTLAL 360  
QY 397 LEVDKHEMSPOAPRRPLPOVATLIDRFVKGSGQLDKDPLDTLTAIFYGYLCSLSPPELS 456  
361 LEVNGHEMS-----POVATLIDRFVKGSGQLDKDPLDTLTAIFYGYLCSLSPPELS 412  
QY 457 SVPPSSIAVARNPOLDTCDPRLDVLVYKARLAFONNMNGSEYFVKIQSFLGAPREDLKA 516  
413 SVPPSSIAVARNPOLDTCDPRLDVLVYKARLAFONNMNGSEYFVKIQSFLGAPREDLKA 472  
QY 517 LSQOVNSMDLATFMKLRDVAVLPLTVAEYOKLGPVHGLAEENHRRYRDMILNOROD 576  
473 LSQOVNSMDLATFMKLRDVAVLPLTVAEYOKLGPVHGLAEENHRRYRDMILNOROD 532  
QY 577 LDTLGLGLOGGIPNGYLVLDLSVOETLSGTPCLLGPVLTVALLLASTLA 628  
533 LDTLGLGLOGGIPNGYLVLDLSVOETLSGTPCLLGPVLTVALLLASTLA 584  
Db

RESULT 5  
US-08-426-819A-37  
Sequence 37, Application US/08426819A  
Patent No. 5723318  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, No. 5723318oml  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihiro  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
TITLE OF INVENTION: Potentiator  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 248 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: Internal  
US-08-426-819A-37

Query Match 38.2%, Score 1247, DB 1, Length 248:  
Best Local Similarity 97.6%, Pred. No. 2,56-111;  
Matches 242; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 37 GETGESAPLGVLTTPHNIISSLSPROLLGFPACAEVSGLSTERYRELAVALAOKNWKST 96  
1 GETGEAPAPLDGVLANPNNISSLSPROLLGFPACAEVSGLSTERYRELAVALAOKNWKST 60  
Db

QY 97 EQLRCLARHLSPEPDDLPDLPLDLFLNPDASFSGPQACTRFFSRITKANVDLPRGAP 156  
61 EQLRCLARHLSPEPDDLPDLPLDLFLNPDASFSGPQACTRFFSRITKANVDLPRGAP 120  
Db

QY 157 RORLLPALACMGVAGSLSEADVAGLACDLPGRVAESAELVLRVSCPEPLD 216  
121 RORLLPALACMGVAGSLSEADVAGLACDLPGRVAESAELVLRVSCPEPLD 180  
Db

QY 217 OOEAAARALOGGPPYGPSTWSVTMDALRGLLPVLGOPITRSIPQIVAMRORSSRD 276  
181 OOEAAARALOGGPPYGPSTWSVTMDALRGLLPVLGOPITRSIPQIVAMRORSSRD 240  
Db

QY 277 PSWROPER 284  
241 PSWROPER 248  
Db

RESULT 6  
US-08-426-819A-33  
Sequence 33, Application US/08426819A  
Patent No. 5723318  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, No. 5723318oml  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihiro  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
TITLE OF INVENTION: Potentiator  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 65 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-426-819A-33

Query Match  
Best Local Similarity 100.0%; Score 325; DB 1; Length 65;  
Best Local Similarity 100.0%; Pred. No. 5.4e-24;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 LIGPCAEVSLSTERYVELVALAOKVKISTEQLRLARLSPPDDALPLDLLF 123  
|||||  
DB 1 LIGPCAEVSLSTERYVELVALAOKVKISTEQLRLARLSPPDDALPLDLLF 60  
|||||

OY 124 LNPPA 128  
|||||  
DB 61 LNPPA 65

RESULT 7  
US-08-426-819A-19  
Sequence 19, Application US/08426819A  
Patent No. 5723318  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, No. 5723318om1  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihiko  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
NUMBER OF INVENTION: Potentilator  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 40 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
FRAGMENT TYPE: internal  
IMMEDIATE SOURCE:  
CLONE: HPC15  
FEATURE:  
NAME/KEY: Peptide

LOCATION: 1..40  
OTHER INFORMATION: /label= fragment  
OTHER INFORMATION: /note= "sequence of cyanogen bromide fragment of MegPOT fr  
OTHER INFORMATION: CDNA in Table 3"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 32  
OTHER INFORMATION: /product= "is Xaa in COS"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 35  
OTHER INFORMATION: /product= "is Xaa in COS"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 37  
OTHER INFORMATION: /product= "is Xaa in COS"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..16  
OTHER INFORMATION: /label= fragment  
OTHER INFORMATION: /note= "sequence of cyanogen bromide fragment representing  
OTHER INFORMATION: C-terminus of MegPOT determined in Example 28"  
US-08-426-819A-19

Query Match  
Best Local Similarity 6.4%; Score 210; DB 1; Length 40;  
Best Local Similarity 100.0%; Pred. No. 2.5e-13;  
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 244 DALRGILPVLCPIRISIPGIVAAWRSSRDPSWROPE 283  
|||||  
DB 1 DALRGILPVLCPIRISIPGIVAAWRSSRDPSWROPE 40  
|||||

RESULT 8  
US-08-426-819A-5  
Sequence 5, Application US/08426819A  
Patent No. 5723318  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, No. 5723318om1  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihiko  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
NUMBER OF INVENTION: Potentilator  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid

STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
FRAGMENT TYPE: Internal  
IMMEDIATE SOURCE:  
CLONE: HPCY5  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..35  
OTHER INFORMATION: /label= peptide  
OTHER INFORMATION: /note= "sequence of a Glu-C peptide of Megpor"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 6..33  
OTHER INFORMATION: /label= peptide  
OTHER INFORMATION: /note= "region of Glu-C peptide having low frequency of usage"  
US-08-426-819A-5

Query Match 5.0%; Score 163; DB 1; Length 35;  
Best Local Similarity 97.1%; Pred. No. 6.4e-09;  
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 83 LAVAIAOKNVKLTSEOLCLARLSEPPEDDLP 117  
DB 1 LAVAIAOKNVKLTSEOLCLARLSEPPEDDLP 35

RESULT 9  
US-09-144-085-2  
Sequence 2, Application US/09144085  
Patent No. 6280999  
GENERAL INFORMATION:  
APPLICANT: Gustafsson, Claes  
APPLICANT: Bellach, Mary C.  
APPLICANT: Ashley, Gary  
APPLICANT: Julien, Bryan  
APPLICANT: Ziermann, Rainer  
TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA  
FILE REFERENCE: 30062-20020.20  
CURRENT APPLICATION NUMBER: US/09/144,085  
CURRENT FILING DATE: 1998-08-31  
EARLIER APPLICATION NUMBER: 09/010,809  
EARLIER FILING DATE: 1998-01-22  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 6095  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-144-085-2

Query Match 4.1%; Score 134; DB 4; Length 6095;  
Best Local Similarity 23.6%; Pred. No. 0.019;  
Matches 143; Conservative 64; Mismatches 217; Indels 182; Gaps 35;

QY 13 GDRPSLLFLFLSGWV--HAKRTLAGETGTSAPLGVLTTPHHISSISPOL----- 64  
DB 2975 GWKPPGTVLITGGVGGIGLHARWLGR-GAEHL-----VLASRCGASAPGASLDELVA 3029  
QY 65 ----LGFCAEYSGLSTERVELAVALAOKNVKLTSEOLCLARL-----RLSE-DPE 111  
DB 3030 RGIRTLACDVS---ERA-QLAALLAE--LEODEAPLRAVAHLAGIGRRVPLELEPE 3082  
QY 112 DLD-----ALPLDLLFLNP-DAF-----SGPOACTRFESRTKANVDL 150  
DB 3083 QLEGLAKVKGAMHLHQHLLGRELDAFVLYGSINGSGQA-----GYGANAGLDAL 3138  
QY 151 PRGAPEROLLPAALACG-----VRSGLSEADVRALGLACDLP-G 192

DB 3139 AR--YRRARGQAATVLMGPMAGEWVTSELESQLRIRGVAVMSPD-KALAGLEMAALRG 3195  
QY 193 RFVASEAEVLPRL-----VSCPPGLDDQDEARAAALQGGPPYGPSTMSVTMDLR 247  
DB 3196 RTSVAIDVDMRSFAPSFSAARPRP-LDGIEARRAESRGP---QPAAGCTALRDTL 3251  
QY 248 GLLPVLGQPIRISIPQGIYAAMRORSSRDPSWROPERTLRFRFRREKTAQPSGKKAR 307  
DB 3252 GLSEAEERRERVQLVASETA--VLGMDP-SRLDPDRGFL----- 3289  
QY 308 EIDESLIFPKKWEACVDAALLATQMDRVNAIPPTY-EOLDVLKHKLDLYPOGYPESV 366  
DB 3290 --DLGDSLMAVELSKRLQK--RTGMYTPSTLSPDHPTQSDVAMWLLLEQLTPQRPBPA 3344  
QY 367 IOHLGVLFKMSPEDIRKKN-----VTSLETKALLEVDKGHEMSPQAR 411  
DB 3345 VR-----EVSREE--GWSPTPIAVGVGLRMPGASDLESMOVL-----VERDYL 3388  
QY 412 RPLP-----QVATLIDRFVKGSGQL--DKDYLDTLTAIFYPGYLGSLSPELSSVPP----- 460  
DB 3389 RPIPAQREVEVLYPPDDPAKKTIVRNASLDDVASFDPGF-GISREAEPMPPQRL 3447  
QY 461 --SSIA-----VRPDLDTCDPRLDVLVYPKARLAFQNMNGSEYFVKIOSFLGAPTE 512  
DB 3448 LLEFMSALEDAGVAPRHLKSGSDTG-----VFVGVAIPSEY-----ASYRGSANE 3492  
QY 513 DLKALS 518  
DB 3493 DAYALT 3498

RESULT 10  
US-09-460-145-2  
Sequence 2, Application US/09460145  
Patent No. 6287838  
GENERAL INFORMATION:  
APPLICANT: Kriz, Ron  
APPLICANT: Song, Chuanzheng  
TITLE OF INVENTION: CYTOSOLIC PHOSPHOLIPASE A2-BETA ENZYMES  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/460,145  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/788,975  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
REFERENCE/DOCKET NUMBER: G15289  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 778 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-460-145-2

Query Match 3.58; Score 113.5; DB 4; Length 778;

Best Local Similarity 20.88; Pred. No. 0.059; Matches 143; Conservative 72; Mismatches 257; Indels 215; Gaps 33;

1 MALORLPCSCGDRPSLFLFLSLGWHPARLAGEGTES---APLG-GVLTTPHNI 56  
 73 MELKVFPODLVTGDDP---VLSVLFDACTLR-----AGEFRESSTLSPOGGRLEVERLR 125  
 57 SSSLPROLLPFPACAEVSGSLSTERYELAVALAOKNVKLSLEQLRCLAHRLSEPPEDDAL 116  
 126 QSLADR-----GEMLVLS---NGVLVARELSCLHYQLLETGDKSSE 163  
 117 PLDLLFLNPDAFSGPQACT-----RFF-SRIKAVNDLPRGAPRQRLPPAALCNG 169  
 164 HRVQLV---PSCGEGPOEASVGTGTFRHCPACWEQELSTIRLODAPEQLKAPLSALPSG 221  
 170 -----VRGSLSEADVR---ALGGLACDLPGRFVAESAELV---LPRLVS 208  
 222 QVAVLVEPTSGEPLMRVLEKKEAGLRRLAVRLGFGPCAEBAFLSRKQVVAALRALQ 281  
 209 CPGLDDOQOABARALOGGPPYGPPTWSVSTMALRGLLPVLGQ---PIRSIPQG 264  
 282 LDGDLQDEIPIVVAIMATGG-----IRAMTSLYGQLAGLKEGLDVCYITGA 331  
 265 IVAMWRORS-SRDSWROPERTILRRFRREVEKTACPSGKKAIDEISLIFFYKMELEA 323  
 332 SGSTWALANLYEDEWSQKD-----LAGPT----- 356  
 324 CVDALLATQMDRVNAPFTYEOLDVLKHLDELYPQGPESVIOHGLYFLKMSPEDIR 383  
 357 ---ELKTVYTKKKLGLVLAISOQRYROELAEARALGYPSCFT----- 396  
 384 KWNVTSLETKALLEVDKGHEMSPOAP---RRPLQVATLIDRFYKRGOLDKDTLDT 438  
 397 ---NLMALINEALLHDEPHDKLSQREALSHGNPLRYCAL-----NTKQO-----S 442  
 439 LTAIFYGLCSLPEELSSVPPSSIMAVRPDLDTC---PROLDVLYPKARLAF-QNNNG 495  
 443 LTFEEFGMCEFSYEVGF---PRYGAFIPSELFGSEFFMQLMKRLPESRICLEGIWS 499  
 496 SEYEVKIQSFL--GGAPTE--DLKALSOQNVSMDLATFMKLR-----TDVVL 538  
 500 NLVYANLQDSLYWASEPQFMDRVNRQANLDKEOVPLKLEPPSTAGRIAEFTDLTLT 559  
 539 --PLTVAEVOKLGRPHVGLKAERHRPRVDMILRORODDLDTGLGLOGGIPN----- 590  
 560 WRPLAQATHNFLRGLHFN--KDYFQHPHFSTW---KATTLQ-----GLPQQLPSE 605  
 591 -----GYVLVDLSVQETLSTPCL 609  
 606 PHCLLDVGYLI-----NTSCL 622

RESULT 11  
 US-09-460-145-4  
 Sequence 4, Application US/09460145  
 Patent No. 6287838

GENERAL INFORMATION:  
 APPLICANT: Song, Chunzheng  
 APPLICANT: Kriz, Ron  
 TITLE OF INVENTION: CYTOSOLIC PHOSPHOLIPASE A2-BETA ENZYMES  
 NUMBER OF SEQUENCES: 9  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genetics Institute, Inc.  
 STREET: 87 Cambridge Park Drive  
 CITY: Cambridge  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02140  
 COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/460,145  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/788,975  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Brown, Scott A.  
 REGISTRATION NUMBER: 32,724  
 REFERENCE/DOCKET NUMBER: G15289  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 498-8224  
 TELEFAX: (617) 876-5851  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 797 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-460-145-4

Query Match 3.58; Score 113.5; DB 4; Length 797;  
 Best Local Similarity 20.88; Pred. No. 0.061; Matches 143; Conservative 72; Mismatches 257; Indels 215; Gaps 33;

1 MALORLPCSCGDRPSLFLFLSLGWHPARLAGEGTES---APLG-GVLTTPHNI 56  
 92 MELKVFPODLVTGDDP---VLSVLFDACTLR-----AGEFRESSTLSPOGGRLEVERLR 144  
 57 SSSLPROLLPFPACAEVSGSLSTERYELAVALAOKNVKLSLEQLRCLAHRLSEPPEDDAL 116  
 145 QSLADR-----GEMLVLS---NGVLVARELSCLHYQLLETGDKSSE 182  
 117 PLDLLFLNPDAFSGPQACT-----RFF-SRIKAVNDLPRGAPRQRLPPAALCNG 169  
 183 HRVQLV---PSCGEGPOEASVGTGTFRHCPACWEQELSTIRLODAPEQLKAPLSALPSG 240  
 170 -----VRGSLSEADVR---ALGGLACDLPGRFVAESAELV---LPRLVS 208  
 241 QVAVLVEPTSGEPLMRVLEKKEAGLRRLAVRLGFGPCAEBAFLSRKQVVAALRALQ 300  
 209 CPGLDDOQOABARALOGGPPYGPPTWSVSTMALRGLLPVLGQ---PIRSIPQG 264  
 301 LDGDLQDEIPIVVAIMATGG-----IRAMTSLYGQLAGLKEGLDVCYITGA 350  
 265 IVAMWRORS-SRDSWROPERTILRRFRREVEKTACPSGKKAIDEISLIFFYKMELEA 323  
 351 SGSTWALANLYEDEWSQKD-----LAGPT----- 375  
 324 CVDALLATQMDRVNAPFTYEOLDVLKHLDELYPQGPESVIOHGLYFLKMSPEDIR 383  
 376 ---ELKTVYTKKKLGLVLAISOQRYROELAEARALGYPSCFT----- 415  
 384 KWNVTSLETKALLEVDKGHEMSPOAP---RRPLQVATLIDRFYKRGOLDKDTLDT 438  
 416 ---NLMALINEALLHDEPHDKLSQREALSHGNPLRYCAL-----NTKQO-----S 461  
 439 LTAIFYGLCSLPEELSSVPPSSIMAVRPDLDTC---PROLDVLYPKARLAF-QNNNG 495  
 462 LTFEEFGMCEFSYEVGF---PRYGAFIPSELFGSEFFMQLMKRLPESRICLEGIWS 518  
 496 SEYEVKIQSFL--GGAPTE--DLKALSOQNVSMDLATFMKLR-----TDVVL 538  
 519 NLVYANLQDSLYWASEPQFMDRVNRQANLDKEOVPLKLEPPSTAGRIAEFTDLTLT 578  
 539 --PLTVAEVOKLGRPHVGLKAERHRPRVDMILRORODDLDTGLGLOGGIPN----- 590

Db 579 WRPLAQAATNHLRGHFR-KDYFOHPHFSTW-KATITD-GLPNQLTPSE 624  
QY 591 -----GYVLDSLVOETLSGTPCL 609  
Db 625 PHCLLDVGYLI-----NTSCL 641

## RESULT 12

US-08-827-208-3  
Sequence 3, Application US/08827208  
Patent No. 6025178  
GENERAL INFORMATION:  
APPLICANT: Chlou, Xue-Chiou C.  
APPLICANT: Kramer, Ruth M.  
APPLICANT: Pickard, Richard T.  
APPLICANT: Sharp, John D.  
APPLICANT: Striffler, Beth A.  
TITLE OF INVENTION: HUMAN PHOSPHOLIPASE A2 AND RELATED  
TITLE OF INVENTION: NUCLEIC ACID COMPOUNDS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/827,208  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/014,608  
FILING DATE: 29-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/041,264  
FILING DATE: 19-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Gaylo, Paul J.  
REGISTRATION NUMBER: 36,808  
REFERENCE/DOCKET NUMBER: X-10610  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317) 276-0756  
TELEFAX: (317) 276-3861  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-827-208-3

Query Match 3.5%; Score 113.5; DB 3; Length 913;  
Best Local Similarity 20.8%; Pred. No. 0.076;  
Matches 143; Conservative 72; Mismatches 257; Indels 215; Gaps 33;

QY 1 MALORLDCPCMGCGRPSGLILFLFSIGVHPARTLAGETGTS--APLG-GVLTTPHNI 56  
Db 208 MELKVFDDDLVTGDDP-VLSVLDAGLR-----AGFFRESSTLSOGGSRLEVERRL 260  
QY 57 SSLSPROLGGFPCAESGLSTERVRELAVALAOKNKLSTEQRLCAHRLSEPPEDDAL 116  
Db 261 QSLADR-----GEMLVS---NGVLVARELSLHVQLEETGQKXSE 298  
QY 117 PLDLLLFPNPAFSGPQACT-----RFF-SRITKANVDLIPGAPERQRLPALACWG 169

Db 299 HRVQLV--PGSCGEPQAEASVGTGTFRHCPACWEQELISRLQDAPEEQALAPLSALPSG 356  
QY 170 -----VRGSLSEADVR-----ALGILACDLPGRFAESAEVL---LPRVLS 208  
Db 357 OVRIVLPPTSOEPLRMVELKREAGLRELAVRIGEGPCAEOAFSRRKOVAAALRLQLQ 416  
QY 209 CGPPLDQDOQEAARAALGGGPRPYGPSTWSVMDALRGLLPVIGQ-----PIRSIPGG 264  
Db 417 LDGDLQDEIDIPVVAIMATGCG-----IRNMTSLYGLAGLKEGLLDCVYITGA 466  
QY 265 IVAAWRORS-SRDPGWRQPERTILPRFRREVEKTCAPSGKAKREIDSLIFYKKMELEA 323  
Db 467 SGSTWALNIYEDPEWSQKD-----LAGPT----- 491  
QY 324 CYDALLATQMDRVAIAFTYEQDLVKKHKLDELTPQGYPPESVYOHLYLFLKMSPEDIR 383  
Db 492 ---ELKTYQVTKNKLGVLPDSQLORYROELAEARLRALGYPSCFT----- 531  
QY 384 KWNYSLETLKALLEVVDKGHEMSPOAP-----RRPLPOVATLIDRFYVGRGOLDKDTLDT 438  
Db 532 --NLWALLNEALLHDEPHDKHLSQREALSHGQNPLPYCAL-----NTKGQ-----S 577  
QY 439 LTAFTPGYLCISPEELSSVPPSSIMAVRPQDLDTCD--PQDVLVYKARLAF-QANNK 495  
Db 578 LTFEFGECFSPYEVGF---PKYGAFIPSELFGESEFFMQLMKRRLPESRICFLEGIMS 634  
QY 496 SEYFVKIOSFL--GGAPTE--DLKALSOQVNSMDLATFMKLR-----TDAYL 538  
Db 635 NLYANLQDSILXMASPEPQFMDRWVRNOANDKEVYPLKLEPPSTAGRIAEFTDLTT 694  
QY 539 --PLTVAEVQKILGPHVEGLKAEERHRPVRWMLRQDDLDITGLGLOGGIPN----- 590  
Db 695 WRPLAQAATNHLRGHFR-KDYFOHPHFSTW-KATITD-GLPNQLTPSE 740  
QY 591 -----GYVLDSLVOETLSGTPCL 609  
Db 741 PHCLLDVGYLI-----NTSCL 757

## RESULT 13

US-09-500-358-3  
Sequence 3, Application US/09500358  
Patent No. 6197569  
GENERAL INFORMATION:  
APPLICANT: Chlou, Xue-Chiou C.  
APPLICANT: Kramer, Ruth M.  
APPLICANT: Pickard, Richard T.  
APPLICANT: Sharp, John D.  
TITLE OF INVENTION: HUMAN PHOSPHOLIPASE A2 AND RELATED  
TITLE OF INVENTION: NUCLEIC ACID COMPOUNDS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/500,358  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/827,208  
FILING DATE: 28-MAR-1997  
APPLICATION NUMBER: US 60/014,608  
FILING DATE: 29-MAR-1996

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/041,264  
FILING DATE: 19-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Gaylo, Paul J.  
REGISTRATION NUMBER: 36,808  
REFERENCE/DOCKET NUMBER: X-10610  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317) 276-0756  
TELEFAX: (317) 276-3861  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-500-358-3

Query Match 3.58; Score 113.5; DB 4; Length 913;  
Best Local Similarity 20.88; Pred. No. 0.076; Indels 215; Gaps 33;  
Matches 143; Conservative 72; Mismatches 257;

QY 1 MALORLPCWSCGDRPSGLFLFLSLGWHPARTIAGETGES---APLG-GVLTTPHNI 56  
DB 208 MELKVFQDOLVTGDDP--VLSVLFDAQTLR-----AGEFRRESFSLSPQSGRLEVERRL 260  
QY 57 SLSLPRLQGLFPFCAEVSGSLSTERYRELAVALAOKNVKISTQNLCLAHRLSEPPEDDAL 116  
DB 261 QSLADR-----GEWLVS---NGVLVAARELSCLHYQLEETDDOKSSE 298  
QY 117 PLDLLFLINPDAFSGPACT-----REF-SRITKANVDLLPGAPERORLLPAAALCWG 169  
DB 299 HRVOLLV--PQSCGPODASVGTGTFRHCPACWQELSTRLQDAPEQLKAPLSALPSG 356  
QY 170 -----VRGSLSEADV---ALGGLACDLPGRFYAESAEVL---LPRLVS 208  
DB 357 QVALVLPPTSOEPLMRVBLKKEAGIRELAVALRGPCAEQAFLSRRKQVAAALRLQALQ 416  
QY 209 CPGLDDQOEAAARALQGGPPPGPSTWVSSTMALRGILLPVLCQ---PIIRSIPOG 264  
DB 417 LDGDLQDEIPVAIMATGG-----IRAMTSLYGOLAGLKEGLDLCVSYITGA 466  
QY 265 IVAAMPORS-SRQSPWROPERTILRPFRPREVEKTACPSGKKAREIDSLIFYKKWLEA 423  
DB 467 SGTWALANLYEDPEWSQKD-----LACPT----- 491  
QY 324 CVDAALLATQMDRVNAIPFTYEOLDLVKHLKLDLYPOGYPESVIQLHGYLFKMSPEDIR 383  
DB 492 -----ELKQVTKRKLGLVLAPOSQLORYRQELAEARALRGYPCFT----- 531  
QY 384 KMWNTSLETALKALEVDKQHEMSPOAP-----RRPLPOVATLIRFYKGRQOLDKDTLDT 438  
DB 532 --NLMALINEALHDEPHDKLSQORALSHGQNPRLPYCAL-----NTKQO-----S 577  
QY 439 LTAAYPGVGLCSPEELSSVPSIMAVRPQDLPTCD--PRQDLVLPKARLAF--QNWNG 495  
DB 578 LITTEFEBCWCFSEYEVGF---PKYGAFFISELGESEPFMQLMKRLPESNICELEGIGWS 634  
QY 496 SEYEVKIQSEFL--GGAPTE--DLKALSOQVNSMDLATFMKLR-----TDAVL 538  
DB 635 NLVYANLQDSLYXMASSEPSQFMDRVNRQANDKQVPLKIEEPPSTAGRIAEFTDLTLT 694  
QY 539 --PLTVAEVOKLLGPHYEGKLAERHRVRWILRQRODDLDITGLGLOGIIPN----- 590  
DB 695 WRPLAQATHNLFRLGHLHN--KDYFOHPHFSTW---KATLTD-----GLPNQLTPSE 740  
QY 591 -----GYLVLDLSVOETLSTPCL 609  
DB 741 PHCLLDVGYLI-----NTSCL 757  
RESULT 14

US-09-498-809-3

Sequence 3, Application US/09498809

Patent No. 6242206

GENERAL INFORMATION:

APPLICANT: Chlou, Xue-Chou C.

APPLICANT: Kramer, Ruth M.

APPLICANT: Pickard, Richard T.

APPLICANT: Sharp, John D.

APPLICANT: Stiffler, Beth A.

TITLE OF INVENTION: HUMAN PHOSPHOLIPASE A2 AND RELATED

TITLE OF INVENTION: NUCLEIC ACID COMPOUNDS

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Eli Lilly and Company

STREET: Lilly Corporate Center

CITY: Indianapolis

STATE: Indiana

COUNTRY: United States of America

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/498,809

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/827,208

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/041,264

FILING DATE: 19-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Gaylo, Paul J.

REGISTRATION NUMBER: 36,808

REFERENCE/DOCKET NUMBER: X-10610

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317) 276-0756

TELEFAX: (317) 276-3861

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 913 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-498-809-3

Query Match 3.58; Score 113.5; DB 4; Length 913;

Best Local Similarity 20.88; Pred. No. 0.076; Indels 215; Gaps 33;

Matches 143; Conservative 72; Mismatches 257;

QY 1 MALORLPCWSCGDRPSGLFLFLSLGWHPARTIAGETGES---APLG-GVLTTPHNI 56  
DB 208 MELKVFQDOLVTGDDP--VLSVLFDAQTLR-----AGEFRRESFSLSPQSGRLEVERRL 260  
QY 57 SLSLPRLQGLFPFCAEVSGSLSTERYRELAVALAOKNVKISTQNLCLAHRLSEPPEDDAL 116  
DB 261 QSLADR-----GEWLVS---NGVLVAARELSCLHYQLEETDDOKSSE 298  
QY 117 PLDLLFLINPDAFSGPACT-----REF-SRITKANVDLLPGAPERORLLPAAALCWG 169  
DB 299 HRVOLLV--PQSCGPODASVGTGTFRHCPACWQELSTRLQDAPEQLKAPLSALPSG 356  
QY 170 -----VRGSLSEADV---ALGGLACDLPGRFYAESAEVL---LPRLVS 208  
DB 357 QVALVLPPTSOEPLMRVBLKKEAGIRELAVALRGPCAEQAFLSRRKQVAAALRLQALQ 416  
QY 209 CPGLDDQOEAAARALQGGPPPGPSTWVSSTMALRGILLPVLCQ---PIIRSIPOG 264  
DB 417 LDGDLQDEIPVAIMATGG-----IRAMTSLYGOLAGLKEGLDLCVSYITGA 466





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 09:48:12 : Search time 143.5 Seconds  
(Without alignments)  
1215.112 Million cell updates/sec

Title: US-09-215-035-2  
Perfect score: 3261  
Sequence: 1 MALGRIDPCWSCGRDRPSLL.....ILGPGVLTVALLLASTLA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

```

1: /cgn2_6/ptodata/2/paa/PCPUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3055	93.7	622	8	US-08-426-819-35
2	2956	90.6	584	8	US-08-426-819-36
3	1547	47.4	328	19	US-09-513-597-2
4	1547	47.4	402	19	US-09-513-597-13
5	1531	46.9	399	9	US-09-513-597-1
6	1247	38.2	248	8	US-08-426-819-37
7	579.5	17.8	175	17	US-09-330-360-733
8	559	17.1	144	17	US-09-330-360-599
9	482	14.8	122	17	US-09-330-360-702

10	424.5	13.0	134	21	US-09-758-458-406	Sequence 406, App
11	325	10.0	65	8	US-08-426-819-33	Sequence 33, Appl
12	220	6.7	144	21	US-09-758-458-404	Sequence 404, App
13	210	6.4	40	8	US-08-426-819-19	Sequence 19, Appl
14	163	5.0	35	8	US-08-426-819-5	Sequence 5, Appl1
15	140	4.3	824	1	PCT-US01-14827-10981	Sequence 10981, A
16	132	4.0	2128	24	US-60-248-505-927	Sequence 927, App
17	131.5	4.0	2088	1	PCT-US01-14827-8424	Sequence 8424, App
18	128.5	3.9	89	22	US-09-874-043-532	Sequence 532, App
19	128.5	3.9	89	22	US-09-875-193-532	Sequence 532, App
20	125	3.8	2114	1	PCT-US01-08631-51615	Sequence 51615, A
21	125	3.8	2114	24	US-60-322-511-970	Sequence 970, App
22	124	3.8	962	21	US-09-758-455-449	Sequence 449, App
23	123.5	3.8	794	18	US-09-456-592-2	Sequence 2, Appl1
24	123.5	3.8	794	21	US-09-789-404-2	Sequence 2, Appl1
25	119	3.6	866	23	US-09-902-540-10786	Sequence 10786, A
26	119	3.6	943	21	US-09-738-626-5199	Sequence 5199, App
27	118	3.6	899	23	US-09-902-540-13735	Sequence 13735, A
28	117.5	3.6	3603	19	US-09-570-581A-1898	Sequence 1898, App
29	117.5	3.6	3603	19	US-09-573-655A-794	Sequence 794, App
30	117	3.6	1415	24	US-60-230-435-1076	Sequence 1076, App
31	117	3.6	1731	21	US-09-739-449-8331	Sequence 8331, App
32	117	3.6	1731	22	US-09-803-110-8331	Sequence 8331, App
33	116	3.6	1429	24	US-60-191-637-35218	Sequence 35218, A
34	116	3.6	1429	24	US-60-191-681-27637	Sequence 27637, A
35	115.5	3.5	842	19	US-09-577-304A-93	Sequence 93, Appl
36	115.5	3.5	842	21	US-09-758-282-93	Sequence 93, Appl
37	114.5	3.5	1016	24	US-60-208-099-22	Sequence 22, Appl
38	114.5	3.5	1016	24	US-60-208-587-35	Sequence 35, Appl
39	114.5	3.5	1310	24	US-60-212-413-205	Sequence 205, App
40	114	3.5	881	16	US-09-352-991A-31702	Sequence 31702, A
41	114	3.5	925	18	US-09-417-507-39237	Sequence 39237, A
42	114	3.5	970	24	US-60-242-679-871	Sequence 871, App
43	114	3.5	1067	24	US-60-161-932-2033	Sequence 2033, App
44	114	3.5	1067	24	US-60-167-217-13306	Sequence 13306, A
45	114	3.5	1067	24	US-60-173-464-10829	Sequence 10829, A

#### ALIGNMENTS

```

RESULT 1
US-08-426-819-35
: Sequence 35, Application US/08426819
:
: GENERAL INFORMATION:
: APPLICANT: Yamaguchi, Nozomi
: APPLICANT: Kojima, Tetsuo
: APPLICANT: Oh-Eda, Masayoshi
: APPLICANT: Hattori, Kunhiro
: TITLE OF INVENTION: Genes Coding for Megakaryocyte
: TITLE OF INVENTION: Potentiator
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: P.O. Box 747
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22040-0747
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/426,819
: FILING DATE: 21-APR-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Murphy Jr., Gerald M.
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 230-107P
: TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 622 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-426-819-35

Query Match 93.7%; Score 3055; DB 8; Length 622;  
Best Local Similarity 95.4%; Pred. No. 1e-275;  
Matches 601; Conservative 3; Mismatches 16; Indels 10; Gaps 3;

QY 1 MALGRDLP-CWSCGDRP-GSLLFLFSLGMYHPARTLAGETGESAIPGLVTPHNIS 58  
DB 1 MALPPARPLDSCGTPALGSLFLFSLGMYQPSRTLAGETGQEAAPLDGVLANPNIS 60  
QY 59 LSPRLGFPCEAVESGLSTERVELAVALAOKNKLSTEQRLAHRLSEPPEDDALPL 118  
DB 61 LSPRLGFPCEAVESGLSTERVELAVALAOKNKLSTEQRLAHRLSEPPEDDALPL 120  
QY 119 DLLFLNPDATSGPOACTRFRSRTKANVDLLPRGAPERQRLPALACWGVRSLSSEA 178  
DB 121 DLLFLNPDATSGPOACTRFRSRTKANVDLLPRGAPERQRLPALACWGVRSLSSEA 180  
QY 179 DVALGGLACDLPGRFVESAIEVLLPRLVSCGPLDDQOEAAARALOGGPPYGPSTW 238  
DB 181 DVALGGLACDLPGRFVESAIEVLLPRLVSCGPLDDQOEAAARALOGGPPYGPSTW 240  
QY 239 SVSTMDALRGLLPVLGQPIIRSIPOGIVAAMRQSSRDPSMRQPERTILRRFRREVEKT 298  
DB 241 SVSTMDALRGLLPVLGQPIIRSIPOGIVAAMRQSSRDPSMRQPERTILRRFRREVEKT 300  
QY 299 ACPGSKKAREIDESLIFKKWELEACVDAALLATQMDRVNAIPFTYEDDLVKHKLDEL 358  
DB 301 ACPGSKKAREIDESLIFKKWELEACVDAALLATQMDRVNAIPFTYEDDLVKHKLDEL 360  
QY 359 POGYPESVIOHLGYFLKMSPEDIRKMNVTSLKLELVKHEMSPOAPRRRLPOVA 418  
DB 361 POGYPESVIOHLGYFLKMSPEDIRKMNVTSLKLELVKHEMSPOAPRRRLPOVA 412  
QY 419 TLIDRFVKGROLDKDTLDTLTAFFPYGLCSLSPBELSSVPPSSITWAVRPDDLTCDPRQ 478  
DB 413 TLIDRFVKGROLDKDTLDTLTAFFPYGLCSLSPBELSSVPPSSITWAVRPDDLTCDPRQ 472  
QY 479 LDVLYPKARLAFQNMNGSEYFVKIQSFLGAPTEDEKALSOQNMMDLATFMKLTDAVL 538  
DB 473 LDVLYPKARLAFQNMNGSEYFVKIQSFLGAPTEDEKALSOQNMMDLATFMKLTDAVL 532  
QY 539 PLTVAEVQKLLGPHVEGKAEERHPRVDWILRQRODDLTGLGLOGGINGYLVLDLS 598  
DB 533 PLTVAEVQKLLGPHVEGKAEERHPRVDWILRQRODDLTGLGLOGGINGYLVLDLS 592  
QY 599 VOETLSTGTCPLGPGPVLTVALALLASTLA 628  
DB 593 VOETLSTGTCPLGPGPVLTVALALLASTLA 622

RESULT 2  
US-08-426-819-36  
Sequence 36, Application US/08426819  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, Nozomi  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hatohri, Kunihiko  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESS: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 584 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
FRAGMENT TYPE: Internal  
US-08-426-819-36

Query Match 90.6%; Score 2956; DB 8; Length 584;  
Best Local Similarity 97.3%; Pred. No. 1.7e-266;  
Matches 576; Conservative 2; Mismatches 6; Indels 8; Gaps 1;

QY 37 GETGTESAPLGGLVLTTPHNISLSLSPROLLGFPCEAVESGLSTERVELAVALAOKNKLST 96  
DB 1 GETGQEAPLDGVLANPNISLSLSPROLLGFPCEAVESGLSTERVELAVALAOKNKLST 60  
QY 97 EQLRLARLSEPPEDDALPLDLILFLNPDATSGPOACTRFRSRTKANVDLLPRGAP 156  
DB 61 EQLRLARLSEPPEDDALPLDLILFLNPDATSGPOACTRFRSRTKANVDLLPRGAP 120  
QY 157 RORLLPALACWGVRSLSSEADYRVALGGLACDLPGRFVESAIEVLLPRLVSCGPLDDQ 216  
DB 121 RORLLPALACWGVRSLSSEADYRVALGGLACDLPGRFVESAIEVLLPRLVSCGPLDDQ 180  
QY 217 QOEAAARALOGGPPYGPSTWSVSTMDALRGLLPVLGQPIIRSIPOGIVAAMRQSSRD 276  
DB 181 QOEAAARALOGGPPYGPSTWSVSTMDALRGLLPVLGQPIIRSIPOGIVAAMRQSSRD 240  
QY 277 PSWRQPERTILRRFRREVEKTACPSGKAREIDESLIFKKWELEACVDAALLATQMDR 336  
DB 241 PSWRQPERTILRRFRREVEKTACPSGKAREIDESLIFKKWELEACVDAALLATQMDR 300  
QY 337 VNAIFTEQDLVKKHKLDELTPGYPESVIOHLGYFLKMSPEDIRKMNVTSLKLEL 396  
DB 301 VNAIFTEQDLVKKHKLDELTPGYPESVIOHLGYFLKMSPEDIRKMNVTSLKLEL 360  
QY 397 LEVDKGHEMSPOAPRRRLPOVATLIDRFVKGROLDKDTLDTLTAFFPYGLCSLSPBELS 456  
DB 361 LEVDKGHEMSPOAPRRRLPOVATLIDRFVKGROLDKDTLDTLTAFFPYGLCSLSPBELS 412  
QY 457 SVPPSSITWAVRPDDLTCDPRQDLVLYPKARLAFQNMNGSEYFVKIQSFLGAPTEDEKAL 516  
DB 413 SVPPSSITWAVRPDDLTCDPRQDLVLYPKARLAFQNMNGSEYFVKIQSFLGAPTEDEKAL 472  
QY 517 LSQQNVSMDLATFMKLTDAVLPLTVAEVQKLLGPHVEGKAEERHPRVDWILRQRODD 576  
DB 473 LSQQNVSMDLATFMKLTDAVLPLTVAEVQKLLGPHVEGKAEERHPRVDWILRQRODD 532

QY 577 LDTGLGLOGGIPNGYLVLDLSVOETLSGTPCLLGGPVLTVLLALLASTLA 628  
|  
Db 533 LDTGLGLOGGIPNGYLVLDLSVOEALSGTPCLLGGPVLTVLLALLASTLA 584

RESULT 3  
US-09-513-597-2

; Sequence 2, Application US/09513597  
; GENERAL INFORMATION:  
; APPLICANT: Scholler, Nathalie B.  
; APPLICANT: Hellstrom, Ingegerd  
; APPLICANT: Hellstrom, Karl Erik  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 730033.410  
; CURRENT APPLICATION NUMBER: US/09/513.597  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-513-597-2

Query Match 47.4%: Score 1547; DB 19; Length 328;  
Best Local Similarity 97.1%: Pred. No. 3.6e-135;  
Matches 301; Conservative 1; Mismatches 0; Indels 8; Gaps 1;

QY 291 FRRVEKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVL 350  
|  
Db 1 FRRVEKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVL 60  
QY 351 KHKIDELYPGSPSVIOHGLGYFLKMSPEDIRKMNVTSLTLLALEVDKGHEMSPOAP 410  
|  
Db 61 KHKIDELYPGSPSVIOHGLGYFLKMSPEDIRKMNVTSLTLLALEVDKGHEMSPOAP 116  
QY 411 RRPPLPVATLIDRFVKGGOGLDKDTLDTLTAFFPGYLCSSPEELSSVPSSIAVVRPOD 470  
|  
Db 117 ---PQVATLIDRFVKGGOGLDKDTLDTLTAFFPGYLCSSPEELSSVPSSIAVVRPOD 172  
QY 471 LDTCDPRDLVLYPKARLAFQNMNGSEYFVXIQSFLGAPTEDEKALSOQVNSMDLATFM 530  
|  
Db 173 LDTCDPRDLVLYPKARLAFQNMNGSEYFVXIQSFLGAPTEDEKALSOQVNSMDLATFM 232  
QY 531 KLRTDAVLP.LTVAEVQKLLGPHVEGLKAEERHRRVDMILRQRODDLTGLGLOGGIPN 590  
|  
Db 233 KLRTDAVLP.LTVAEVQKLLGPHVEGLKAEERHRRVDMILRQRODDLTGLGLOGGIPN 292  
QY 591 GYLVLDLSVQ 600  
|  
Db 293 GYLVLDLSVQ 302

RESULT 4  
US-09-513-597-13

; Sequence 13, Application US/09513597  
; GENERAL INFORMATION:  
; APPLICANT: Scholler, Nathalie B.  
; APPLICANT: Hellstrom, Ingegerd  
; APPLICANT: Hellstrom, Karl Erik  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 730033.410  
; CURRENT APPLICATION NUMBER: US/09/513.597  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 402  
; TYPE: PRT  
; ORGANISM: Homo sapien

US-09-513-597-13

Query Match 47.4%: Score 1547; DB 19; Length 402;  
Best Local Similarity 97.1%: Pred. No. 5.1e-135;  
Matches 301; Conservative 1; Mismatches 0; Indels 8; Gaps 1;

QY 291 FRRVEKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVL 350  
|  
Db 1 FRRVEKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVL 60  
QY 351 KHKIDELYPGSPSVIOHGLGYFLKMSPEDIRKMNVTSLTLLALEVDKGHEMSPOAP 410  
|  
Db 61 KHKIDELYPGSPSVIOHGLGYFLKMSPEDIRKMNVTSLTLLALEVDKGHEMSPOAP 116  
QY 411 RRPPLPVATLIDRFVKGGOGLDKDTLDTLTAFFPGYLCSSPEELSSVPSSIAVVRPOD 470  
|  
Db 117 ---PQVATLIDRFVKGGOGLDKDTLDTLTAFFPGYLCSSPEELSSVPSSIAVVRPOD 172  
QY 471 LDTCDPRDLVLYPKARLAFQNMNGSEYFVXIQSFLGAPTEDEKALSOQVNSMDLATFM 530  
|  
Db 173 LDTCDPRDLVLYPKARLAFQNMNGSEYFVXIQSFLGAPTEDEKALSOQVNSMDLATFM 232  
QY 531 KLRTDAVLP.LTVAEVQKLLGPHVEGLKAEERHRRVDMILRQRODDLTGLGLOGGIPN 590  
|  
Db 233 KLRTDAVLP.LTVAEVQKLLGPHVEGLKAEERHRRVDMILRQRODDLTGLGLOGGIPN 292  
QY 591 GYLVLDLSVQ 600  
|  
Db 293 GYLVLDLSVQ 302

RESULT 5  
US-09-513-597-1

; Sequence 1, Application US/09513597  
; GENERAL INFORMATION:  
; APPLICANT: Scholler, Nathalie B.  
; APPLICANT: Hellstrom, Ingegerd  
; APPLICANT: Hellstrom, Karl Erik  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 730033.410  
; CURRENT APPLICATION NUMBER: US/09/513.597  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 399  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-513-597-1

Query Match 46.9%: Score 1531; DB 19; Length 399;  
Best Local Similarity 97.1%: Pred. No. 1.6e-133;  
Matches 298; Conservative 1; Mismatches 0; Indels 8; Gaps 1;

QY 294 EVEKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVLKHK 353  
|  
Db 1 EVEKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVLKHK 60  
QY 354 LDELYPGSPSVIOHGLGYFLKMSPEDIRKMNVTSLTLLALEVDKGHEMSPOAPRRP 413  
|  
Db 61 LDELYPGSPSVIOHGLGYFLKMSPEDIRKMNVTSLTLLALEVDKGHEMSPOAPRRP 113  
QY 414 LPQVATLIDRFVKGGOGLDKDTLDTLTAFFPGYLCSSPEELSSVPSSIAVVRPOD 473  
|  
Db 114 -PQVATLIDRFVKGGOGLDKDTLDTLTAFFPGYLCSSPEELSSVPSSIAVVRPOD 172  
QY 474 CDPRLDLVLYPKARLAFQNMNGSEYFVXIQSFLGAPTEDEKALSOQVNSMDLATFMKLR 533  
|  
Db 173 CDPRLDLVLYPKARLAFQNMNGSEYFVXIQSFLGAPTEDEKALSOQVNSMDLATFMKLR 232

QY 534 TDAVLPLTVAEYOKLLGPHVEGLKAEERHPRVDILNRQRODDLTGLGLOGGIPNGYL 593  
|||||  
Db 233 TDAVLPLTVAEYOKLLGPHVEGLKAEERHPRVDILNRQRODDLTGLGLOGGIPNGYL 292  
QY 594 VLDLSVQ 600  
|||||  
Db 293 VLDLSVQ 299

## RESULT 6

US-08-426-819-37  
; Sequence 37, Application US/08426819  
; GENERAL INFORMATION:  
; APPLICANT: Yamaguchi, Nozomi  
; APPLICANT: Kojima, Tetsuo  
; APPLICANT: Oh-Eda, Masayoshi  
; APPLICANT: Hattori, Kunihiko  
; TITLE OF INVENTION: Genes Coding for Megakaryocyte  
; TITLE OF INVENTION: Potentiator  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/426,819  
; FILING DATE: 21-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 230-107P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-205-8000  
; TELEFAX: 703-205-8050  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEetical: NO  
; FRAGMENT TYPE: Internal  
; US-08-426-819-37

Query Match 38.2%; Score 1247; DB 8; Length 248;  
Best Local Similarity 97.6%; Pred. No. 2.6e-107;  
Matches 242; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 37 GETGTESAPLGGVLTTPNINISLSRQLLGFPCAESVLSLTERVELAVLAOKNVKIST 96  
|||||  
Db 1 GETGTESAPLGGVLTTPNINISLSRQLLGFPCAESVLSLTERVELAVLAOKNVKIST 60  
QY 97 EOLRLARLSEPPEDDALPLDILLFLNPDAFSGPOACTRFESRITKANVDLPRGAPE 156  
|||||  
Db 61 EOLRLARLSEPPEDDALPLDILLFLNPDAFSGPOACTRFESRITKANVDLPRGAPE 120  
QY 157 RQRLPALACMGVSGSLSEADVRAALGCLDLGRFVAESAELPLRLVSCPPPLDOD 216  
|||||  
Db 121 RQRLPALACMGVSGSLSEADVRAALGCLDLGRFVAESAELPLRLVSCPPPLDOD 180  
QY 217 QOEAAARALGGGPPYGPSTWSVTMDALRGLPLVLCOPITIRISPOGIVAAARORSSRD 276  
|||||

Db 181 QOEAAARALGGGPPYGPSTWSVTMDALRGLPLVLCOPITIRISPOGIVAAARORSSRD 240  
QY 277 PSMROPER 284  
|||||  
Db 241 PSMROPER 248

## RESULT 7

US-09-330-360-733  
; Sequence 733, Application US/09330360  
; GENERAL INFORMATION:  
; APPLICANT: Gearing, David P.  
; APPLICANT: Robison, Keith E.  
; APPLICANT: Holtzman, Douglas A.  
; TITLE OF INVENTION: Nucleic Acid Molecules Derived from a  
; TITLE OF INVENTION: Human Fetal Lung Library  
; FILE REFERENCE: MLN98-34PA  
; CURRENT APPLICATION NUMBER: US/09/330,360  
; CURRENT FILING DATE: 1999-06-11  
; PRIOR APPLICATION NUMBER: 60/090,258  
; PRIOR FILING DATE: 1998-06-22  
; NUMBER OF SEQ ID NOS: 1102  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO: 733  
; LENGTH: 175  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)..(33)  
; NAME/KEY: VARIANT  
; LOCATION: (1)..(175)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
; US-09-330-360-733

Query Match 17.8%; Score 579.5; DB 17; Length 175;  
Best Local Similarity 68.6%; Pred. No. 3.8e-45;  
Matches 129; Conservative 5; Mismatches 23; Indels 31; Gaps 5;

QY 1 MALQRLDP-CWSCGRP-GSILFLFSIGWHPARTLAGGCTESAPLGGVLTTPNINISS 58  
|||||  
Db 1 MALPAPRLGSCGTPALGSLLEFLFSIGWQPSHTLAGTGQEAAPLDGVLANPNISS 60  
QY 59 LSPRLGFPCAESVSGSLTERVELAVLAOKNVKISTEOLRCLAHRLSEPPEDDALPL 118  
|||||  
Db 61 LSPRLGFPCAESVSGSLTERVELAVLAOKNVKISTEOLRCLAHRLSEPPEDDALPL 120  
QY 119 DLLFLNPDAFSGPOACTRFESRITKANVDLPRGAPERORLL-----PALAC- 167  
|||||  
Db 121 DLLFLNPDAFSGPOA-----XPVSSPASRRPMWTCSESGSRATAAAGC 165  
QY 168 ---WGVRG 172  
|||||  
Db 166 SGLMGVVG 173

## RESULT 8

US-09-330-360-599  
; Sequence 599, Application US/09330360  
; GENERAL INFORMATION:  
; APPLICANT: Gearing, David P.  
; APPLICANT: Robison, Keith E.  
; APPLICANT: Holtzman, Douglas A.  
; TITLE OF INVENTION: Nucleic Acid Molecules Derived from a  
; TITLE OF INVENTION: Human Fetal Lung Library  
; FILE REFERENCE: MLN98-34PA  
; CURRENT APPLICATION NUMBER: US/09/330,360  
; CURRENT FILING DATE: 1999-06-11  
; PRIOR APPLICATION NUMBER: 60/090,258  
; PRIOR FILING DATE: 1998-06-22  
; NUMBER OF SEQ ID NOS: 1102  
; SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 599  
LENGTH: 144  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: (1)...(33)  
NAME/KEY: VARIANT  
LOCATION: (1)...(144)  
OTHER INFORMATION: Xaa - Any Amino Acid  
US-09-330-360-599

Query Match 17.1%; Score 559; DB 17; Length 144;  
Best Local Similarity 86.0%; Pred. No. 2.2e-43;  
Matches 117; Conservative 2; Mismatches 15; Indels 2; Gaps 2;

QY 1 MALORLDP-CWSCGDRP-GSLFLFLFSLGWHPARTLAGETTESAPLGGVLTTPHNIS 58  
DB 1 MALPTARFLGSCGTPALGSLFLFLFSLGWQPSRTLAGETGQEAAPLDGVLANPNISS 60  
QY 59 LSPROLLGFPCEAVSGSLTERVRELAVLAQKNVLSLEQLCLAHRLSEPPEDIDALPL 118  
DB 61 LSPROLLGFPCEAVSGSLTERVRELAVLAQKNVLSLEQLCLAHRLSEPPEDIDALPL 120  
QY 119 DLLFLNPDAFSGPOA 134  
DB 121 DLLFLNPDAFSGPOA 136

RESULT 9  
US-09-330-360-702  
Sequence 702, Application US/09330360  
GENERAL INFORMATION:  
APPLICANT: Gearing, David P.  
APPLICANT: Robison, Keith E.  
APPLICANT: Holtzman, Douglas A.  
TITLE OF INVENTION: Nucleic Acid Molecules Derived from a  
FILE REFERENCE: M1N98-34PA  
CURRENT APPLICATION NUMBER: US/09/330,360  
CURRENT FILING DATE: 1999-06-11  
PRIOR APPLICATION NUMBER: 60/090,258  
PRIOR FILING DATE: 1998-06-22  
NUMBER OF SEQ ID NOS: 1102  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 702  
LENGTH: 122  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: (1)...(33)  
US-09-330-360-702

Query Match 14.8%; Score 482; DB 17; Length 122;  
Best Local Similarity 84.3%; Pred. No. 2.7e-36;  
Matches 102; Conservative 2; Mismatches 15; Indels 2; Gaps 2;

QY 1 MALORLDP-CWSCGDRP-GSLFLFLFSLGWHPARTLAGETTESAPLGGVLTTPHNIS 58  
DB 1 MALPTARFLGSCGTPALGSLFLFLFSLGWQPSRTLAGETGQEAAPLDGVLANPNISS 60  
QY 59 LSPROLLGFPCEAVSGSLTERVRELAVLAQKNVLSLEQLCLAHRLSEPPEDIDALPL 118  
DB 61 LSPROLLGFPCEAVSGSLTERVRELAVLAQKNVLSLEQLCLAHRLSEPPEDIDALPL 120  
QY 119 D 119  
DB 121 D 121

RESULT 10  
US-09-758-458-406  
Sequence 406, Application US/09758458  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PM041  
CURRENT APPLICATION NUMBER: US/09/758,458  
CURRENT FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
NUMBER OF SEQ ID NOS: 520  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 406  
LENGTH: 134  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (51)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (56)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (97)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (117)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (127)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (131)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-758-458-406

Query Match 13.0%; Score 424.5; DB 21; Length 134;  
Best Local Similarity 63.7%; Pred. No. 7.5e-31;  
Matches 93; Conservative 3; Mismatches 23; Indels 27; Gaps 3;

QY 263 QGIVAAARORSSRDSPWRPERTILRRPRREVEKTACSGKARIDSLFFYKMELE 322  
DB 7 RSGGKWR-----RQPVLODEGPRDRREPH-----LQEVGAG----- 39  
QY 323 ACVDALLATOMDRVNAIPFTYEOLDKLHKLDELXPQGPSPSVIOHGLFLKNSPEDI 382  
DB 40 TCVDALLATQXDRVNAIPFTYEOLDKLHKLDELXPQGPSPSVIOHGLFLKNSPEDI 99  
QY 383 RKWNTSLETLKALLEVDKGHMSPO 408  
DB 100 RKWNTSLETLKALLEVDKGHMSPO 125

RESULT 11  
US-08-426-819-33  
Sequence 33, Application US/08426819  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, Nozomi  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihiko  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
TITLE OF INVENTION: Potentiator  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church

STATE: Virginia  
 COUNTRY: USA  
 ZIP: 22040-0747  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/426,819  
 FILING DATE: 21-APR-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murphy Jr., Gerald M.  
 REGISTRATION NUMBER: 28,977  
 REFERENCE/DOCKET NUMBER: 230-107P  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703-205-8000  
 TELEFAX: 703-205-8050  
 INFORMATION FOR SEQ ID NO: 33:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 65 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-426-819-33

Query Match 10.0%; Score 325; DB 8; Length 65;  
 Best Local Similarity 100.0%; Pred. No. 4,5e-22;  
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 LIGFPCAEEVSGSTERVELAVALAOKNKVLTSEOLRCIAHRLSEPPEDLDLPDLILF 123  
 DB 1 LIGFPCAEEVSGSTERVELAVALAOKNKVLTSEOLRCIAHRLSEPPEDLDLPDLILF 60  
 QY 124 LNPDA 128  
 DB 61 LNPDA 65

RESULT 12  
 US-09-758-458-404  
 Sequence 404, Application us/09758458  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PM041  
 CURRENT APPLICATION NUMBER: US/09/758,458  
 CURRENT FILING DATE: 2001-01-11  
 PRIOR APPLICATION NUMBER: 60/179,065  
 PRIOR FILING DATE: 2000-01-31  
 PRIOR APPLICATION NUMBER: 60/180,628  
 PRIOR FILING DATE: 2000-02-04  
 NUMBER OF SEQ ID NOS: 520  
 SOFTWARE: Patentin Ver. 2.0  
 SEQ ID NO 404  
 LENGTH: 144  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (128)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 NAME/KEY: SITE  
 LOCATION: (137)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 NAME/KEY: SITE  
 LOCATION: (140)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 US-09-758-458-404

Query Match 6.7%; Score 220; DB 21; Length 144;  
 Best Local Similarity 47.5%; Pred. No. 1,1e-11;  
 Matches 58; Conservative 3; Mismatches 11; Indels 50; Gaps 5;

QY 4 QRLDPCW-----SCG-----DRPSLLFLFLSLGWHPARTLGE 38  
 DB 25 QRLDPCWGVGPSPASASCSASDSCSPRGPMLEQGRR-----HPWTDLA-- 73  
 QY 39 TGTESAPLGAVLTPTNINISLSPROLLGFPCEVSGSLSTERVELAVALAOKNVK-LSTE 97  
 DB 74 -----NPNISLSPROLLGFPCEVSGSLSTERVELAVALAOKNVQALQPE 120  
 QY 98 QL 99  
 DB 121 QI 122

RESULT 13  
 US-08-426-819-19  
 Sequence 19, Application US/08426819  
 GENERAL INFORMATION:  
 APPLICANT: Yamaguchi, Nozomi  
 APPLICANT: Kojima, Tetsuo  
 APPLICANT: Oh-Eda, Masayoshi  
 APPLICANT: Hattori, Kunihiko  
 TITLE OF INVENTION: Genes Coding for Megakaryocyte  
 NUMBER OF SEQUENCES: 37  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Birch, Stewart, Kolasch & Birch  
 STREET: P.O. Box 747  
 CITY: Falls Church  
 STATE: Virginia  
 COUNTRY: USA  
 ZIP: 22040-0747  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/426,819  
 FILING DATE: 21-APR-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murphy Jr., Gerald M.  
 REGISTRATION NUMBER: 28,977  
 REFERENCE/DOCKET NUMBER: 230-107P  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703-205-8000  
 TELEFAX: 703-205-8050  
 INFORMATION FOR SEQ ID NO: 19:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 40 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: peptide  
 HYPOTHETICAL: NO  
 FRAGMENT TYPE: Internal  
 IMMEDIATE SOURCE:  
 CLONE: HPCY5  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 1..40  
 OTHER INFORMATION: /label= fragment  
 OTHER INFORMATION: /note="sequence of cyanogen bromide fragment of MegPOT fr  
 OTHER INFORMATION: CDNA in Table 3"  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: 32  
 OTHER INFORMATION: /product="is Xaa in COS"  
 FEATURE:

```
NAME/KEY: Modified-site
LOCATION: 35
OTHER INFORMATION: /product= "is Xaa in COS"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 37
OTHER INFORMATION: /product= "is Xaa in COS"
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..16
OTHER INFORMATION: /label= fragment
OTHER INFORMATION: /note= "sequence of cyanogen bromide fragment representing
OTHER INFORMATION: C-terminus of MegPOT determined in Example 28"
US-08-426-819-19

Query Match      6.48; Score 210; DB 8; Length 40;
Best Local Similarity 100.0%; Pred. No. 1.1e-11;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 244 DALRGLPVIGPPIIRSIPOGIVAAWRORSRSDPSWROPE 283
Db 1 DALRGLPVIGPPIIRSIPOGIVAAWRORSRSDPSWROPE 40

RESULT 14
US-08-426-819-5
Sequence 5, Application US/08426819
GENERAL INFORMATION:
APPLICANT: Yamaguchi, Nozomi
APPLICANT: Kojima, Tetsuo
APPLICANT: Oh-Eda, Masayoshi
APPLICANT: Hatcori, Kunihiko
TITLE OF INVENTION: Genes Coding for Megakaryocyte
TITLE OF INVENTION: Potentiator
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESS: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,819
FILING DATE: 21-Apr-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 230-107P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 35 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHEtical: NO
FRAGMENT TYPE: Internal
IMMEDIATE SOURCE:
CLONE: HPCY5
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..35
```

```
OTHER INFORMATION: /label= peptide
OTHER INFORMATION: /note= "sequence of a Glu-C peptide of MegPOT"
FEATURE:
NAME/KEY: Peptide
LOCATION: 6..33
OTHER INFORMATION: /label= peptide
OTHER INFORMATION: /note= "region of Glu-C peptide having low frequency of us
OTHER INFORMATION: codons."
US-08-426-819-5

Query Match      5.08; Score 163; DB 8; Length 35;
Best Local Similarity 97.1%; Pred. No. 2.3e-07;
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 83 LAVAALAKNVKLTSTEOQLAHRLSEPPEDDALP 117
Db 1 LAVAALAKNVKLTSTEOQLAHRLSEPPEDDALP 35

RESULT 15
PCT-US01-14827-10981
Sequence 10981, Application PC/TUS0114827
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-104
CURRENT APPLICATION NUMBER: PCT/US01/14827
CURRENT FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 09/577,408
PRIOR FILING DATE: 2000-05-18
NUMBER OF SEQ ID NOS: 16102
SOFTWARE: Custom
SEQ ID NO 10981
LENGTH: 824
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (747)..(768)
OTHER INFORMATION: POLYCYSTIC KIDNEY DISEASE PROTEIN SIGNATURE domain identified
OTHER INFORMATION: by eMATRIX, accession number PR00500B, p-value=4.316e-09, raw
PCT-US01-14827-10981

Query Match      4.38; Score 140; DB 1; Length 824;
Best Local Similarity 23.38; Pred. No. 0.0063;
Matches 155; Conservative 70; Mismatches 267; Indels 172; Gaps 32;

Oy 15 RPSGL--LFLFSLGWHPARTLAGETGESAPLGVLTPPHN-ISSLSPPQLLGFPQAE 71
Db 194 RPTFGALGALATLNLNHNALVYLRPMAFQGLLVYRRLSLNHNALSVLRPALNGLPALR 253
Oy 72 VSGLSTERELAVALAOKNVKLTSTEDLCIAH-RLSEPP-----EDLDALP--LDLL 122
Db 254 RLSHHNELQALPGV-----LSQARGLARLEIGHNPVTYAGEEDGLALPGLEELL 305
Oy 123 -----FLNPAFSGPQACTREFSRITKAN--VDLLP-----RGAPRQRLLAALAC----- 167
Db 306 DGAALQALGPRAFA--HCPRLHTLDLRGNQDLPPLQGGQLRRLRLGNNPLMCGQA 362
Oy 168 -----MGVRSLSLEADVRALGGLACDLPGRFVAESAELVLRPLVSCPGPDDOQEAR 222
Db 363 RPLEW-----LARAVRSDG--ACGPRRLRGEALDALRPMDLRCGDAAQEEEELEE 414
Oy 223 AALQG-----GPPYGPSTWSVTMDLRGLPLVIGPI-----INSIPOGIVAAWROR 272
Db 415 RAVAGPRAPRPGPGRGPEEBRAVAPCPRACVCPESRHSCEGGLQAVPRGF----- 467
Oy 273 SSRPSMRQPERTILRPFRREVEKTACPSGKKAREIDSLIFKKMPLKLCVDAALLAT 332
Db 468 ---PS--DTQLDLRRNHFPSVPRAPF-----GLGHVLSLHLHGIAIELEAG 511
```



```

OY 333 QMDRVNAIPFTY---EOLDVLKHKLDELYPQGYPESVIOHGLYFLKMSPEDIRKWNVTS 389
Db 512 ALAGLGRLIYLYLSDNOLAGLSAALLEGAPR-----LGYLYERN----- 551
OY 390 LETTKALLEVDKGHMSPOAPRRPLQVATLLIDRFVKGROLDKDLDTLTAFYPGYL-- 447
Db 552 -----RFLQV-----PGAALRALPSLSL-----HLQDNADVRLA---PGDLGR 587
OY 448 -----CSLSPSELSSVPSSIMAVRPQDDLTCDPRQDLVLYPKARLAFQNMNGSEYFV 500
Db 588 TRALRMGLHLGNNRTTEVSLGALGPARELEKLIHLDNOLREVPTGA-----LEGIPALL 641
OY 501 KIOSEFGAPTEDEKALSQQWNSMDLATEFMKLRDVAVLPLTVAEVOKLIGPHEGL--KA 558
Db 642 ELQ--LSGNPLRALBDGAFQVYGRSLQ-----LFLNSSGLEQYGTGHLAGLYQEA 690
OY 559 EERRRPVRDWILRORODDLTGLGLGGIPNGYLVLDLSVOETLSGTPCLLGGPVLTV 618
Db 691 AQGH-----RORAFYQQAFAFASPLVPGIGLQSLHLQ-KNQLRALPAL-----PSLSQ 737
OY 619 LALL 622
Db 738 LELI 741

```

Search completed: December 7, 2001, 09:52:20  
 Job time: 248 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 09:48:02 ; Search time 26.14 Seconds  
(without alignments)  
244.470 Million cell updates/sec

Title: US-09-215-035-2  
Perfect score: 3261  
Sequence: 1 MALQRIDPCWCGDRPGSL.....LIGPFLVTLALLASTLA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 57046 seqs, 10175881 residues

Total number of hits satisfying chosen parameters: 57046

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	579.5	17.8	175	US-09-828-792-733	Sequence 733, App
2	559	17.1	144	US-09-828-792-599	Sequence 599, App
3	482	14.8	122	US-09-828-792-702	Sequence 702, App
4	131.5	4.0	6	US-60-325-795-2	Sequence 2, Appli
5	102	3.1	1159	US-09-815-242-11089	Sequence 11089, A
6	101.5	3.1	2472	US-09-815-242-5064	Sequence 5064, Ap
7	100	3.1	956	US-09-815-242-11925	Sequence 11925, A
8	99.5	3.1	941	US-09-722-487-55	Sequence 55, Appli
9	99.5	3.1	941	US-09-722-708-55	Sequence 55, Appli
10	99.5	3.1	3739	US-09-657-440-2	Sequence 2, Appli
11	99	3.0	610	US-60-325-795-4	Sequence 4, Appli
12	96.5	3.0	308	US-09-976-594-449	Sequence 188, App
13	96.5	3.0	308	US-09-981-353-188	Sequence 188, App
14	95.5	2.9	689	US-09-897-516-6629	Sequence 6629, Ap
15	95.5	2.9	4342	US-09-815-242-5107	Sequence 5107, Ap
16	95	2.9	1062	US-09-815-242-5111	Sequence 5111, Ap
17	94.5	2.9	1048	US-09-815-242-10062	Sequence 10062, A
18	93	2.9	667	US-09-815-242-12811	Sequence 12811, A
19	93	2.9	3816	US-09-808-880-3	Sequence 3, Appli
20	92.5	2.8	937	US-09-964-899-51	Sequence 51, Appli
21	92	2.8	353	US-09-897-516-8074	Sequence 8074, Ap
22	92	2.8	657	US-09-815-242-5408	Sequence 5408, Ap
23	92	2.8	667	US-09-815-242-12490	Sequence 12490, A
24	92	2.8	1527	US-09-897-516-4688	Sequence 4688, Ap
25	91.5	2.8	1656	US-09-897-516-6555	Sequence 6555, Ap
26	91	2.8	349	US-09-919-935-2	Sequence 2, Appli
27	90	2.8	451	US-09-964-899-31	Sequence 31, Appli

28	90	2.8	519	US-09-897-516-6004	Sequence 6004, Ap
29	89.5	2.7	9799	US-09-965-738-146	Sequence 146, App
30	89.5	2.7	11721	US-09-965-738-162	Sequence 162, App
31	89	2.7	1115	US-09-915-543-17	Sequence 17, Appli
32	88.5	2.7	200	US-09-815-242-5053	Sequence 5053, Ap
33	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
34	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
35	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
36	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
37	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
38	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
39	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
40	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
41	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
42	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
43	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
44	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
45	88.5	2.7	775	US-09-989-723-326	Sequence 326, App

ALIGNMENTS

RESULT 1  
US-09-828-792-733  
Sequence 733, Application US/09828792  
GENERAL INFORMATION:  
APPLICANT: Gearling, David P.  
APPLICANT: Robison, Keith E.  
APPLICANT: Holtzman, Douglas A.  
TITLE OF INVENTION: Nucleic Acid Molecules Derived from a  
FILE REFERENCE: MLN98-34pa  
CURRENT APPLICATION NUMBER: US/09/828,792  
CURRENT FILING DATE: 2001-04-09  
PRIOR APPLICATION NUMBER: US/09/330,360  
PRIOR FILING DATE: 1999-06-11  
PRIOR APPLICATION NUMBER: 60/090,258  
PRIOR FILING DATE: 1998-06-22  
NUMBER OF SEQ ID NOS: 1102  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 733  
LENGTH: 175  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: (1)...(33)  
NAME/KEY: VARIANT  
LOCATION: (1)...(175)  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-828-792-733

Query Match 17.8%; Score 579.5; DB 5; Length 175;  
Best Local Similarity 68.6%; Pred. No. 2.4e-42;  
Matches 129; Conservative 5; Mismatches 23; Indels 31; Gaps 5;  
QY 1 MALQRIDPCWCGDRPGSLFLFLFSIGWVHPARTLAGETESAPLGVLTTPHNSS 58  
DB 1 MALPFARLSCGCPALGSLFLFLFSIGWVQPSSTLAGEGQEAAPDGVLANPNNSS 60  
QY 59 LSPROLLGFCFAEVSGLSTERYRELAVLAQKNVKTLEQLRCLAHRLSEPPEDDALPL 118  
DB 61 LSPROLLGFCFAEVSGLSTERYRELAVLAQKNVKTLEQLRCLAHRLSEPPEDDALPL 120  
QY 119 DLLFLNPDAFSGQACTRFRSRTTKANVDDLPRCAERORL-----PAALAC- 167  
DB 121 DLLFLNPDAFSGQACTRFRSRTTKANVDDLPRCAERORL-----PAALAC- 165  
QY 168 ---MGVRC 172  
DB 166 SGLMGVRC 173



```
OY 390 LETLKALLEVDKHEMSPQAPRRPLPOVATL-----IDREFVKG 427
Db 508 -----RFLQV-----PGALRALPSLFSLHLDONAVDRLAPG 539

RESULT 5
US-09-815-242-11089
: Sequence 11089, Application US/09815242
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA 011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 11089
: LENGTH: 1159
: TYPE: PRT
: ORGANISM: Haemophilus influenzae
US-09-815-242-11089

Query Match 3.18; Score 102; DB 5; Length 1159;
Best Local Similarity 20.28; Pred. No. 1.7;
Matches 149; Conservative 96; Mismatches 272; Indels 222; Gaps 34;

OY 29 VHPARTLAGEGTSTAPLPGVLTTPHNTSSLSPPR-----OLLGFPCA 70
Db 495 INMARKLEGVTRMKNKAGVVISPTLITDFAPLYCDNEGLHPVTHPKNDVEYAGLYKF 554
OY 71 EVSGLSREVELAVALLAKNVKLTSTPOLRCIARHRLSEPPELDALPLDLLFLNPDASF 130
Db 555 DELGRLTTLTKMLDIT--NVBMYRE-----GKPRVDIAIPLD-----DPSEFE 598
OY 131 --GPOACTREFSRITKANVDLLPFCAPRORLLPALALCMGVRGSLSEADVALGGLAC 188
Db 599 LLKRSETTAVFQLSRRGKMDIKRLQDP-----CF-----EDIAL--VAL 637
OY 189 DLPRFVAESAEVLLPRL-----VSCPRPLDQDOGEA--RAALGGGPGPYPRPSTWSVT 242
Db 638 FRPGFLOGVWDFNLDNRHGREEVSY--DAEYOHASLKLPLE--PTYG-----682
OY 243 MDALRGLLPVLOPIRISIPQIYAAV-----RORSSRDPSPWRPPEFTILPRFR 293
Db 683 -----IILYQEVVMIOAQ--VLAGTYLGADLLRRAMGKKKPEPMAKORLVEGEGAK 733
OY 294 E-----VEKTACPSGKKAREIDSLFYK---KWELEACVDALLATQMDR 336
Db 734 NGIDGELSMKIFDLVEKTFAGYGFMKSHSAAYALVSYQTLMLKTHPBAFMAAVMTSEMDN 793
```

```
OY 337 VNAIPETVEQ-----LDVLKHLDELVPOG---YPESVIOHGX-----372
Db 794 TDKIVGLYDECLRMGLKATPPDINIGKHNF--SVNEQGEIVYIGAIKVGCGSPIEALVAA 852
OY 373 -----LFLKMSPEDIRKMNVTSETTLKALLEVDKHEMSPQ--APRRPLPOVATLI 421
Db 853 RNEGCIKFDLFDLCARVLDKIKNRTFESLIMSCAFDR--LGPFRALSLNLEDAALRAS 909
OY 422 DREYKRGQLDKRLDILT-----AFYPGYLC--LSPE-----LSSVPPSI 463
Db 910 DOHAKDEAMGQTDMEGVLTETHEDVENAVANTPYTEKOIILDGERETGLYLSHPVSRY 969
OY 464 WAVRPODLDTCDDPOLDVLYPKARLAFQNMNGSEFVYIQ-----SFLGAPTEDLKALS 518
Db 970 L-----KELSHYSTRKLKLANRRGOISTVAGLVYASNIAMTKGNRLGATLDD-----1020
OY 519 QQNVSMDLATF-----MKLRTDVAV-----LPLTYAEVOKL--LGP 552
Db 1021 -RSGRDLTLFGESLDEQGEKLDQDTVIIVAGQVYFDDFSGGLKMTVRELMTLDEARSRY 1079
OY 553 VEGKAEERHPRVDMILRQDODLD-----TLGLGLOGIIPNGYVLVDSVOETLSGP 607
Db 1080 VKSLAISLSEHQITPSPFKQFKALPEVSGGLTLPINYYQSPKGRALRLGOWSIIPT- 1138
OY 608 CLLGPGVLTVALILLAST 626
Db 1139 -----DELTELVLNLGES 1152

RESULT 6
US-09-815-242-5064
: Sequence 5064, Application US/09815242
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA 011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5064
: LENGTH: 2472
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5064

Query Match 3.18; Score 101.5; DB 5; Length 2472;
Best Local Similarity 21.79; Pred. No. 5.6;
Matches 138; Conservative 70; Mismatches 216; Indels 211; Gaps 33;
```



ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/722.487  
FILING DATE: 28-Nov-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/179,558  
FILING DATE: <Unknown>  
APPLICATION NUMBER: U.S. 60/063,898  
FILING DATE: 31-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 9426-005-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 941 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 55:  
US-09-722-487-55

Query Match 3.1%; Score 99.5; DB 5; Length 941;  
Best Local Similarity 20.8%; Pred. No. 2;  
Matches 148; Conservative 83; Mismatches 216; Indels 263; Gaps 40;

58 SLSPROLLGPPCAVSGSTERVELAY---ALAKNVKLTSTEQRLCIAHRLSEPPEDL 113  
2 SISP---LKPCCL---LQSRGKKIEANRQKALARAEXKLADQHOKRPAQSKGSPSNTL 54  
114 DALPLDL---LRF-----LNPDAF-----SGPOACTRF 138  
55 PRDSKSGSHCIFKQONPSSSHGDGRPNHSFPPTSTQAKGMQRPPEMPTACPSY 114  
139 F--SRITKANVDL---LPRGAPERORLLPRAALACWVRSGLLSBADVRALGGLACDL-P 191  
115 RPPMOVTVAGISLPLANSPPGVPSQQ-----LWG-----CELGQ 148  
192 GRFAESAENVLPRLVSCGFLPDQDOEAARALQGGPPYCP-----PST-----WS 239  
149 GHQASLETOSTPFRANTTHERLKRKYKNQETFAASSSGOPRPPDELEARIIGRPSTSGNIS 208  
240 VSTMDALRGLLP-----VLGQPIITRSIPQG-----IVANRQSSR--D 276  
209 GSVMPRTGRLQOKAGTRPMHNVGSSOGRICRNGERFQVKIGYNEALIAVFKSLPSRSYD 268  
277 P---SW-----ROBERTILRRPFRRE--VEKTACPSCKKAREIDESLIFY 316  
269 PATTMNFSMTDVPPLKKAORLPGLT-LQPLEGAEGHMESSPSTSSG-----TI 316  
317 KKWELACVDAALLATQMDRVNAIPFTYEQDLVKHKLDELDELPOGYPEPVTOHLYGLFLK 376  
317 AKTGLPAPASLAFYKGCQVLSSRARF---EADI-----SYSDLLA---LFGQ 358  
377 MSPDIDKMNVT-----LETALAILEVDKGHEMSPOAPRRRPLPQVATLIDRVPYKGRGQ 430

Db 359 M---DSRKDYVKTRKMSFLLEEYSKLMERVNG---PPQVOLDPLPKTLTLF-----RAQ 406  
Qy 431 LDKDTLDTLTAFPEYGLCSLSPEELSSVPSISMAVRQDDTDQDPROLDVLYP-----484  
Db 407 LQK-----TSLSP--VADLPEA-----DISRVDKSLVSSLLPFORAGV 442  
Qy 485 -----KARLAFQUNNG-----SEYFVKIOSFLGAPTED-----LKAL 517  
Db 443 NFATAQRGRLLADMDGIGKTIQAICTIAATYRKEMPLLVVPPSSVFTWEOAFCHMLPSL 502  
Qy 518 SQQVNSMDLAFEMKLRDVAVPLT---VAEVOKLLGPHVEGLAEERHRRPVRDMLLRQ 573  
Db 503 NPLDIN-VVTKGRDLRTDGLVINVSFDLSKLEOKLPKFKVVIIDESH-----550  
Qy 574 QDDDLTGLGLQGGIPNGIYVLDLSVOE-TISGIPCLLGPVULT-VLAL 621  
Db 551 --FLKNITAVCAAMP---LLKVARVILLSGTPAMSPAEVLTQIILAV 594

## RESULT 9

US-09-722-708-55

Sequence 55, Application US/09722708  
GENERAL INFORMATION:  
APPLICANT: Hockensmith, Joel W.  
Muthuswami, Robin

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
TARGETING DNA METABOLIC PROCESSES USING  
AMINOGLYCOSIDE DERIVATIVES

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESSES:

ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/722,708

FILING DATE: 28-Nov-2000

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/179,558

FILING DATE: &lt;Unknown&gt;

APPLICATION NUMBER: U.S. 60/063,898

FILING DATE: 31-OCT-1997

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 9426-005-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)7909090

TELEFAX: (212)8699741

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 941 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 55:  
US-09-722-708-55

Query Match 3.1%; Score 99.5; DB 5; Length 941;  
Best Local Similarity 20.8%; Pred. No. 2;  
Matches 148; Conservative 83; Mismatches 216; Indels 263; Gaps 40;

```

OY      58 SLSPHOLLGFPCEAEVSGJSTRVELAV-----ALAAOKVVKJSTQOLCLAHRISEPEDL 113
Db      2 SISP--LKCCCL--LQSRGKKIENRKKALARAEREKLLAQHQKPAOSKOGPQONL 54
OY      114 DALPLD-----LLF-----LMPDAF-----SGPQACTRF 138
Db      55 PRDPKSGSHGIFFEKOONPSSSSHGDRPDPNDHSPFPPTSEQAKGMQOREEMPTTACPSY 114
OY      139 F--SMITKANLD-----LPRAPERORLLPRALACMVGVRGSLSEADVRALGGLACL-P 191
Db      115 RPNQVYTAGISLPLANSRPGVPSSQ-----LMG-----CELGO 148
OY      192 GRVAESAEEVLLPRLVSCRPRLDDQOEAAALAQGGGPPYGP-----PSF-----WS 239
Db      149 GHQASLETSQSTPRANTHTERLKRKYKNQNETRAASSSQGPRDPRLERIGRPSISGNIS 208
OY      240 VSTDALGGLP-----VLGQPIRISPGC-----IYAANRQSSR--D 276
Db      209 GSVMPRTGRLQOKAGRPMPHRVSVSOGKCRFRNEROVYKIGYNEALIAVFKSILSPRSYD 268
OY      277 P--SM-----ROBERTILRPRFRRE--VEKTACPSGKKAREIDESLIFY 316
Db      269 PATKWNNSMTDYGRLKAAORLEGIT-LQPLBEGAHMESPSTSSG-----II 316
OY      317 KKEELACVDALLATOMDRVNAIPFTYEQDLVHLKHLDELPGGYRPSYIQLGYLFLK 376
Db      317 AKTGIPAPSLAFVYGQCVLISRRARF--EADI-----SYSEDLIA--LFKQ 358
OY      377 MSPEDIRKWNYS-----LEFTKALLEYQKGHMSPOAPRRPLQVAVATLIDRVKRGQ 430
Db      359 M--DSKRYIDVKTKRKWSLLEYSKLEWRNG--PROVOLDRPLPKTLTP--RAQ 406
OY      431 LDKDTLDTLTAFFPGYGLCSLSPHEELSVSPSSIAVNRPRODITCDPQOLDVLYP----- 484
Db      407 LQK-----TSLSP--VADIPEA-----DLSRVDSKLYVSSLPLPFRQAGV 442
OY      445 -----KRLAFQNNNG-----SEYVKIQSFLGCAPTD-----LKAL 517
Db      443 NFALAQRRLLADMDMGKTQAICTAAAYRKEMPLLIVVPSSVRFTEWQAFCRMPLSL 502
OY      518 SQQVNSMDLAFMKLTDAVPLT-----VAEVOKLGPHEVGLAEERHRPVRWILQR 573
Db      503 NPLOIN-VVIGKRLRDGLVNIYSPDLSSKLEQOLKRPFRVYIIDSH----- 550
OY      574 QDDLDITGLGQGGIPNGYVLVLSVOE-TLSGPPCLLGPVLT-VLAL 621
Db      551 --FLKNITAVCAAMP-----LLKVAKRVIILSGTPAMSRPAELTYQIILAV 594

RESULT 10
US-09-657-440-2
Sequence 2, Application US/09657440
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BEPLACH, Melanie C.
APPLICANT: BETFLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, LI
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/657,440
PRIOR APPLICATION NUMBER: 09/330,878
PRIOR FILING DATE: 1999-05-27
PRIOR APPLICATION NUMBER: CIP OF 09/141,908
PRIOR FILING DATE: 1998-08-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 3739
TYPE: PRF
ORGANISM: Streptomyces venezuelae
US-09-657-440-2

```

Db 14 LPLVLALLPAAQAAARCPACICDNRHRYACRYONLTPVDAIPELTQRL----- 67  
QY 167 CMGVRGSLSDADYRAGLA-----CDLPGRFPAESA-----EVLPLRVSCGP 212  
Db 68 --DLGNLKVTPAARAFQGVPHLTHLDLRCEV--ELVABGAFGLGRLLNLNAS--NH 121  
QY 213 LDDOQDEA-----AANALOGGPPYGPSTW-----SVSTMDALRGILPVLAGPIIRSI 261  
Db 122 LRELPGALDGLSLRLEEGNALFEELRPGTFCALGALATLNLHNLVYL--PAMAF- 178  
QY 262 POGIV-AAAROSSRSDPSWROPERTILRPRFR-----EVEKACPSGKAR-----EI 309  
Db 179 -OGILRRWMLSHNALSVLAEALAGLPALRRLSLHNEIQALPGVLSQARGIAREL 237  
QY 310 DESLIFYKME-----LEACVDALLATQMDRVNAIPFTYEQDLYLKHKLDELVP-Q 360  
Db 238 GHNLVTAGEEDGLALPGRLLELDGALQALGPRAPRHCPRCLHTLDRGNQDLPPLQ 297  
QY 361 GYPSVIOHLG-----YFLKMSPEDIRKMNVTSETLKL--LEVDKGHE 404  
Db 298 GPGGLHQHGCIAGLEAGALAGLGRLLIYILSDNQLAGLSAALGAPRLGIVLERNR 357  
QY 405 MS-PQAPRRPLQVATL-----IDREYKG 427  
Db 358 LQVGAALRALPSLSLHLDQNAVDRAPG 387

## RESULT 12

US-09-976-594-449  
; Sequence 449, Application US/09976594  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT FILING DATE: 2001-10-12  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 449  
; LENGTH: 308  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: 522433CD1  
US-09-976-594-449

Query Match 3.0%; Score 96.5; DB 5; Length 308;

Best Local Similarity 22.9%; Pred. No. 0.69; Mismatches 126; Indels 137; Gaps 19;

Db 19 LPLFLSLGWNHPARTLAGETGSAPLGVLTPPHNITSLSPROLGFPCEAVSGLSTE 78  
QY 14 MLVLVLVSWL-----PHGAL-----SLAEASRASFP--GSESLHTE 49  
Db 79 --RVRELAVALAOKNVKISTEOLRCLAHRLSEPPEDDALPLDLILFLNPAFSGPOACT 136  
QY 50 DSRRELKRRYEDLLTRLRANQ-----SMEDSNTDLVPAPAVRILTPVRLGSGG-- 99  
Db 137 RFRSRITKANVDLPKGRPERQ-----RLPAAACWGRGSLSDADYRAGLGACD 189  
QY 100 HLHLRISRRA--ALPEGLPEASRLHRLFLSPPTASRSDVTRPLRLQLSL----- 147  
Db 190 LPRGFVAESAVALPRLVSCGPLDDO-----QEARA-----ALOG 227  
QY 148 -----ARQAPALHLRL--SEPPSGSDQLLAESSARPOLLEHLRPOAKRRRARANG 200  
QY 228 GGPYGPSTSVSTMDALRGILPVLAGPIIRSIPOGIVAAAROSSRSDPSWROPERTIL 287

Db 201 DHCPGRCRCRLHTVRA-----SLEDLGMAD---WVL 230  
QY 288 RPRRREVEKT-----ACSGSKAREIDESLIFYKKWELEACVDALLATQMDRVNA---I 340  
Db 231 SP---REVQVVMCIAGACSOFRRA-----NMHAQIKTSLHRLKPDVTPAPCCV 275  
QY 341 PFTYEQLDVLKHKLD 355  
Db 276 PASYNPM-VLIQKTD 289

## RESULT 13

US-09-981-353-188  
; Sequence 188, Application US/09981353  
; GENERAL INFORMATION:  
; APPLICANT: Laasek, Amy W.  
; APPLICANT: Jones, David A.  
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
; FILE REFERENCE: PA-0038 US  
; CURRENT APPLICATION NUMBER: US/09/981,353  
; CURRENT FILING DATE: 2001-10-11  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PERL Program  
; SEQ ID NO 188  
; LENGTH: 308  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: 522433CD1  
US-09-981-353-188

Query Match 3.0%; Score 96.5; DB 5; Length 308;

Best Local Similarity 22.9%; Pred. No. 0.69; Mismatches 126; Indels 137; Gaps 19;

QY 19 LPLFLSLGWNHPARTLAGETGSAPLGVLTPPHNITSLSPROLGFPCEAVSGLSTE 78  
Db 14 MLVLVLVSWL-----PHGAL-----SLAEASRASFP--GSESLHTE 49  
QY 79 --RVRELAVALAOKNVKISTEOLRCLAHRLSEPPEDDALPLDLILFLNPAFSGPOACT 136  
Db 50 DSRRELKRRYEDLLTRLRANQ-----SMEDSNTDLVPAPAVRILTPVRLGSGG-- 99  
QY 137 RFRSRITKANVDLPKGRPERQ-----RLPAAACWGRGSLSDADYRAGLGACD 189  
Db 100 HLHLRISRRA--ALPEGLPEASRLHRLFLSPPTASRSDVTRPLRLQLSL----- 147  
QY 190 LPRGFVAESAVALPRLVSCGPLDDO-----QEARA-----ALOG 227  
Db 148 -----ARQAPALHLRL--SEPPSGSDQLLAESSARPOLLEHLRPOAKRRRARANG 200  
QY 228 GGPYGPSTSVSTMDALRGILPVLAGPIIRSIPOGIVAAAROSSRSDPSWROPERTIL 287  
Db 201 DHCPGRCRCRLHTVRA-----SLEDLGMAD---WVL 230  
QY 288 RPRRREVEKT-----ACSGSKAREIDESLIFYKKWELEACVDALLATQMDRVNA---I 340  
Db 231 SP---REVQVVMCIAGACSOFRRA-----NMHAQIKTSLHRLKPDVTPAPCCV 275  
QY 341 PFTYEQLDVLKHKLD 355  
Db 276 PASYNPM-VLIQKTD 289

## RESULT 14

US-09-897-516-6629  
; Sequence 6629, Application US/09897516  
; GENERAL INFORMATION:  
; APPLICANT: Cordin, David R.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.



APPLICANT: Huesing, Joseph E.  
APPLICANT: Krasomil-Osterfeld, Karina C.  
APPLICANT: Malvar, Thomas M.  
APPLICANT: Slater, Steven C.  
APPLICANT: Spiridonov, Sergei  
TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof  
FILE REFERENCE: 38-21(51847)B  
CURRENT APPLICATION NUMBER: US/09/897,516  
CURRENT FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/215, 161  
PRIOR FILING DATE: 2000-06-30  
NUMBER OF SEQ ID NOS: 8409  
SEQ ID NO 6629  
LENGTH: 689  
TYPE: PR  
ORGANISM: Xenorhabdus sp.  
US-09-897-516-6629

Query Match 2.9%; Score 95.5; DB 5; Length 689;  
Best Local Similarity 23.4%; Pred. No. 2.8;  
Matches 91; Conservative 47; Mismatches 130; Indels 121; Gaps 21;

QY 44 APPLG---VITPHNISLSPROLGFPCAESGL---STERYRELAVALAOKNVL 95  
DB 88 AGLGGLEFALACHARICSLDKTRGLPEVOLGLPGSGGTORLRLIGVSSALDMILT 147  
QY 96 TEQJRC-LAHLSEPPEDLALPLDLFL-----NPDASF-----GPQA 134  
DB 148 GRLKAKAKAKLGYVD---DAVPLDILDVAVOYKKVAKRKPLAMSORLLASSFGRL 204  
QY 135 CTREFSRITKAN-----VDLPRGAPERORLPAALACWGVGSLSEADVAL 183  
DB 205 LFRWQGTAKTRKTHGYHAPERIIDIVIRLLEKGGK---AGL-----QAEKAF 250  
QY 184 GGLACD---LPGFVA-----ESAIVLPRLVSCGPELDODQOEARALOGGEPY 232  
DB 251 GELMWSSSSALRLFFFTTSIKNETGSELPAOTIRHYGILG-----GGLMGCG--- 299  
QY 223 GPPSTWSTMDALRGLPVLGPI-IRSIPOGIYAAWRQSSR-----DPSMRQPERT 285  
DB 300 -----IANTATRGHLPAKIKDINKEGINQALKYTMDMSKRSORLIPSESRQMS 352  
QY 286 ILPRFRREVKTACPSGKAREI---DESLIFYKKW---ELEACVDAALLATQMDRVN 338  
DB 353 LL-----SGTIDYSGEQADIVYEAIVEDLYLKRKVAEVEANTKAETIFA--SMTS 402  
QY 339 AIPFTYEQDLVKKHLDLYPOGYPEYI 367  
DB 403 SLPI-----HOIAETIAKR--PEOVI 420

RESULT 15  
US-09-815-242-5107  
Sequence 5107, Application US/09815242  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5107  
LENGTH: 4342  
TYPE: PR  
ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-5107

Query Match 2.9%; Score 95.5; DB 5; Length 4342;  
Best Local Similarity 22.2%; Pred. No. 42;  
Matches 89; Conservative 61; Mismatches 149; Indels 101; Gaps 21;

QY 214 DODQOEARALOGGPGPYGPPSTWSTMDALRGLPVLGPIITRSIPGIVAAWRORS 273  
DB 1579 EAEQORLRAALQAEPLPYWPT---OLMRLAQWPLPGSKLPTRALPEPY---WOORE 1631  
QY 274 SRDPS-----WRQP---ERTILPRF-----RREVEKTACPSGKAREID 310  
DB 1632 HVEPTEILQRRIAAIWSEVLGLPRVGLRDPFELGSHSLATRIYSRI---RACDVE 1686  
QY 311 ESL-IFYKKWELEA-C--VDALALA-----TOMDRVAIPFTYEQDLVKKHKLDEL 357  
DB 1687 LPLRALFEASELEACEGVRAQAAGRTDSHCALIRIDREQPVPLSYSQ---QRMVFL 1741  
QY 358 YPOGYPEVIOHLYLFLKMSPEIDIRKNNVTSLETLKALLVEYDKGHEMSPOAPRRPLQY 417  
DB 1742 W-OLEPDSPAYNVGGLARLSGPLDVARFEA---ALQALVO---RHE---TLRTTPEVS 1789  
QY 418 ATLIDRFYKGGOLDKDTLDTLTAFFYPGYLSLSPSELSSVPSSIMAVRPDDTCDPR 477  
DB 1790 DGVPQVRHVGCGGLHMDQD-----FSALDRDSRQOHLQTLADSE--AHRPFDES--- 1838  
QY 478 QLDVLYPRARLAFOMMNGSEYFVKIOSFLGAPTEDELKALSQOQVNSMDL-----AT 528  
DB 1839 -----GPLLRCOMYKMAERHYLVY-----TLHHIYTEGWAMDIFARELGALEYA 1883  
QY 529 FMKLRITDAVLPITYAEVOKLGLPHVEGILKAEHRIRPVQW 568  
DB 1884 FLDRRESPLEPLPVQYLDYSWQR-EMLESGEROROLDYW 1922

Search completed: December 7, 2001, 09:49:36  
Job time: 94 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 09:52:27 : Search time 33.02 Seconds  
(without alignments)  
1448.747 Million cell updates/sec

Title: US-09-215-035-2  
Perfect score: 3261  
Sequence: 1 MALQRDLPCWCGSDRPSGL.....LTCGPVLTVALLASTLA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR\_68:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1708	52.4	625	2 JC7362	Erc protein - rat
2	125.5	3.8	2109	2 T31352	hypothetical prote
3	120	3.7	2218	2 B84683	hypothetical prote
4	111.5	3.4	4957	2 T03455	ALR protein - huma
5	111.5	3.4	5262	2 T03454	ALR protein - huma
6	110.5	3.4	340	2 B96632	hypothetical prote
7	110	3.4	1770	2 T18551	salivarycn Mx1 syn
8	109	3.3	1171	2 T35548	hypothetical prote
9	108.5	3.3	652	2 G82962	hypothetical prote
10	108.5	3.3	1397	2 T51292	Dna2p - fission ye
11	108.5	3.3	1398	2 T39568	hypothetical helic
12	108.5	3.3	2361	2 T25752	hypothetical prote
13	108	3.3	283	2 E82662	ABC transporter AT
14	107.5	3.3	1288	2 T42756	5-oxoprolinase (AT
15	107.5	3.3	3573	2 S23070	erythronolide synt
16	106.5	3.3	653	2 A46362	amyloid precursor
17	106.5	3.3	1748	1 JN0786	integrin beta-4 ch
18	106	3.3	391	2 T36311	probable lipase -
19	106	3.3	527	2 E75503	probable fibronect
20	106	3.3	690	2 S35251	probable membrane
21	106	3.3	1121	2 A82809	exodeoxyribonuclea
22	105.5	3.2	1607	2 T13250	hypothetical prote
23	105	3.2	617	2 F75484	serine/threonine-s
24	105	3.2	693	1 A41090	ankyrin-like prote
25	105	3.2	830	2 E83031	probable metallopr
26	105	3.2	1058	2 D82654	hypothetical prote
27	105	3.2	1096	2 T08619	hypothetical prote
28	105	3.2	2298	2 T49648	hypothetical prote
29	104.5	3.2	264	2 S75053	hypothetical prote

30	104	3.2	503	2 A83027	hypothetical prote
31	104	3.2	1003	2 B71469	glycine--trna lig
32	104	3.2	7576	2 T17428	FK506 polyketide s
33	103.5	3.2	637	2 A75342	hypothetical prote
34	103.5	3.2	1402	2 S75938	chemotaxis protein
35	103.5	3.2	4128	2 JC6306	protein kinase (EC
36	102.5	3.1	436	2 T28066	hypothetical prote
37	102.5	3.1	733	2 H69411	cell division cont
38	102.5	3.1	1010	2 T36383	probable large ATP
39	102.5	3.1	1239	2 G71266	probable ATP-depen
40	102	3.1	1159	1 H64089	DNA-directed DNA p
41	102	3.1	1160	2 S34809	avrBs3-2 protein -
42	101.5	3.1	836	2 T30312	plin biosynthetic
43	101.5	3.1	2472	2 E83594	still frame shift p
44	101	3.1	677	2 G02540	nucleobindin - hum
45	101	3.1	1997	2 T30874	virgInIamycin S sy

## ALIGNMENTS

RESULT 1  
JC7362  
Erc protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 31-Dec-2000  
C:Accession: JC7362  
R:Yamashita, Y.; Yokoyama, M.; Kobayashi, E.; Takai, S.; Hino, O.  
Biochem. Biophys. Res. Commun. 275, 134-140, 2000  
A:Title: Mapping and determination of the cdna sequence of the Erc gene preferentiall  
A:Reference number: JC7362  
A:Accession: JC7362  
A:Molecule type: mRNA  
A:Residues: 1-625 <YAM>  
A:Cross-references: DDBJ:D87351  
C:Comment: This protein, as a tumor marker for renal cell carcinoma, is involved in c  
C:Genetics:  
A:Gene: Erc  
A:Map position: 10q12

Query Match	52.4%	Score 1708;	DB 2;	Length 625;
Best Local Similarity	55.7%	Pred. No. 3.1e-114;		
Matches 351;	Conservative 81;	Mismatches 184;	Indels 14;	Gaps 5;
OY	1	MALQRDLPCWCGSDR--PGSILFLFLSLGVNHPARTLAGETGSAPLGGVLTTPPHNIS	57	
DB	1	MALPTAQLPLGSCGSPICSRSLFLLLSLISGLWPLRLQOTTRTSQEAALLHAVTGTV-DFA	59	
OY	58	SLSPPQLLGFPCAEEVSGSLSTERVRELAVAAOKNVKLSLEOLRCIAHRLSE--PREDDA	115	
DB	60	SLPTLFLGLTCDDEVSGLSMGNAKELAMAVKOKNVLYOVHQRCLARLRPKLITLNEEDDA	119	
OY	116	LPDLFLFLNPDAPFSGPOACTRFESRITKANVDLLPAGAPERORLLPALACWGVRSGLL	175	
DB	120	LPDLFLFLNPFMPGQQAACHFFSLISKANVNVLPRLSLEDORQLTLGALKQGVGYGV	179	
OY	176	SEADVRALGGLACDLPGRFVAESAELLPLRVSCGCPDLDDQQAARALDGGGPPYGP	235	
DB	180	SETDARALGGLACDLPGEFVAKSSSEVLLPWLARCGLPDGQAKAVREVLRSGRAPYGP	239	
OY	236	STWASITMDALRGLLPVIGOPILIRSIPOGIVAAWQRSSRDPSPQPERTILRPFRRREV	295	
DB	240	STWASITMDALRGLLVLDSEIVHSIPKDVITTEWLQGISREPSRSGSKWTYHHPFRKDT	299	
OY	296	EKTACPSGKKAKEIDSLFYKKWLEACVDALLATQMDRYNAIPTYEQDLVLKHLKD	355	
DB	300	EOKACPPGKEPVDNDENLIFYQNMELACVDTLLAGOMDLVNEIPFTYEQLSIFKHKLD	359	
OY	356	ELYPGYPESVYQHGIYFLFKMSPDIDKAVNTSLETETKALLEVVGKGMESQAPRRRLP	415	
DB	360	KTYPGYPESLTIKQGHFRRVYSPEDIRQWNTSPDVTNLTLSKSGKQMD-----A	411	

QY 416 QVALIDREFKRGROLDKOTLDTLTAIFYEYICLSLSPPEELSSVPPSSIAAVRPOLDLTD 473

Db 412 QVALVACETYLRGGKIDEDIVKALDNIPIUSYLCDFSPDLMHIPSVMVLVELHDLDKCS 471

QY 476 PROLDVLYPARLAFQNMNGSEYFVKIOSFLGAPREDIKALSOOVNSDMLATPFMKLRTD 535

Db 472 QRHGIILYKACASAFQVWSLEYFEKIRITPLGASNSEDIKALSOHNVSMDITFFKMOYD 531

QY 536 AVLPVTAEVQKLLGPHVESLKAEERHRRPDMILRODDLDITGLGIGGIPNGYIVL 595

Db 532 ALVGLSAEYQKELGPHIGDLKTEEDKSPVDMILFRQOQKDDLSGLGIGGIPNGYITL 591

QY 596 DLASQELTSGPCILGPGPLVTALILLAS 625

Db 592 DENVREAFSSGAPDLGGPFVFAITALLAS 621

```

RESULT      2
T31352
hypothetical protein - Pelargonium x hortorum
C:Species: Pelargonium x hortorum
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31352
R:Downie, S.R.; Katz-Downie, D.S.; Wolfe, K.H.; Calie, P.J.; Palmer, J.D.
Curr. Genet. 25, 367-378, 1994
A:Title: Structure and evolution of the largest chloroplast gene (ORF2280): Internal plast
A:Reference number: 221012; MUID:94363755
A:Accession: T31352
A:Status: preliminary; translated from GR/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-2108 <DOM>
A:Cross-references: EMBL:M83200; NID:9466913; PID:9468914; PIDN:AAAT3173.1

```

Query Match	3.8%;	Score 125.5;	DB 2;	Length 2109;
Best Local Similarity	20.5%;	Pred. No. 2.2;		
Matches 128;	Conservative 91;	Mismatches 207;	Indels 199;	Gaps 32

QY	79	RVRRELVAAQKNVKNLSTEQUR-----CLARRSEPPEDDIALPDLLEFNPDPAFSGPQ	133
Db	1006	RIREIS-IRMCIRNLTLSERIRNNESPDLTTHLRSPNVLVEFLYSTLLLL-----VAGYL	1060
QY	134	ACTREFSRITKANVDL-----LPRGAPRORLL-----PAALACMGVRSLSSEA	178
Db	1061	VCT-YLRSLSKDYELQTELKVKVSLMIPSTYTELRKIMDRYPSELMSFELKMLFLVAM	1119
QY	179	DVRAIGGIACDLPGRFVAESAENVLLPRIVSCPGDLDDQGEAARAALQGGCPGP---	235
Db	1120	EE-----LKESLSVYGM-----LEGGRATGVEAIF	1146
QY	236	STMSVSTMDALRGLLPYVGOPITRSIPQIVAAVRQNRSDPSMRQPERTLLRFRFRREV	295
Db	1147	SNMNLNID-ISDLISLIPNID-ITFSITRHLSHTSKE-----IYSITRRER---	1195
QY	296	EKTACPSGKKAREIDESILIFKKW---ELEACVAAALLATMDRVNAIPF-----TYEQ	346
Db	1196	-----YGANIDDKIESLSTSAIDDCDGNLQGFSTLLITTEKG	1236
QY	347	LDVLKHLDELPGQPESVYOHILGYLFIKMSPPEDIRK-----WNVTSLFETLKALLE-	398
Db	1237	VDOILLSLTQSSKAASSQOMEQPEMYLR-HVVDLQKTKYIMGYEPNTSSLAERRIFLAH	1295
QY	399	-----VDKGHEMSFOAP---RRPL--POVATLIDRFVKGSGROLDKD-----	434
Db	1296	YQMTYTSKTSQGVNAHFHPSHEKPFSSFRDLSPRGILVLISGTGNSYLKLSLAKTNHF	1355
QY	435	-----TL-----DITLAFYPGYLSPELSVSPSSITWAVRPODLL	472
Db	1356	PLITLEMARSSMPEFQHLDEITYGQLEUYVDDPSLSEVBE-----BEDTSWGIEWMSLP	1411
QY	473	TCDPROLDVLYPKARL-AFQNMNGSEYFVKIQSLGCAFTEDLKALSOQVNSMDLATFMK	531
Db	1412	-JTRREDELEDOAQEMQTRRDDIEYTHAIODMI-----DL-----GISVAAQ--LN	1455

```

OY 532 LRIAYVLTVAEQKLGSHVEGLKAEEENRRVYRWMLRQKQDDLDLGLGLDGG-----587
Db 1456 FLPESL-----INEGHVDVDEALEL-----WELMEEMDRHLGLISLPAALVRA 1505
OY 588 -----IPNGVLYLDSVQETLSG 605
Db 1506 MTPCIIIMIPNHDV-DLEERTTAG 1529

```

### RESULT 3

hypothetical protein At2g28300 [mouse] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84683  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii,  
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Umayam, L.; Talbot,  
Euss, D.; Niemann, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: AB4420, MIMD:20083487  
A:Accession: B84683  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2218 (STO)  
A:Cross-references: GB:AE002093; NID:g4803953; PIDD:AAD29825.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g28300  
A:Map position: 2

Query Match 3.7%; Score 120; DB 2; Length 2218;  
Best Local Similarity 19.1%; Pred. No. 5.9;  
Matches 135; Conservative 81; Mismatches 222; Indels 270; Gaps 32.

QY	30	HPARTLA-----GEMTESAPLGVLTTPHNITSUSPROLGGFPCAEWSGSLTERVRE	82
Db	169	HPTSLATISPDLSGPPFOSLP-----ASPAPTPIRGRGRGRGRGAGGR	217
QY	83	LAVALAOKNVKLS--TBQLRCLARHLSPPDDLDLFLPNPAFGSPQACTFFS	140
Db	218	VEGVHGSNNSITORTETNSLA-----SDAEA-----TKFALPRASIEIVS	259
QY	141	RITKANVDLPRGAPER-ORLLPALACMCVRSGLSEADYRALGLACLPGRFEVESA	199
Db	260	RVPKAN-----EGSTSNPDQYSPVHSATTLARSKADKDL-----DAPREFDSGH	306
QY	200	EVLLPRLVSCGPLDDOQEA-----RAALGGGPPYGPSPWSV-----STMD	244
Db	307	-----VQTLNVLENSSEKKAFAVKKRRLPLGGGPVQNONNAVSVCDSKSPSEGRTYT	359
QY	245	ALREL-----LPVLGPI-----IRSTPGIYAAMRQR---	272
Db	360	ALQGVTTAPSDATLPMSSOPSDATLPMSSOPVGSTVEAOEANPSPALPAKRRVNL	419
QY	273	SSRPSMNRPER-----TII-----RPRFR	293
Db	420	SRGETPRKGRKROPLRPAYDASSARSTGLTPOLEVYVNLGCTKAKEDAVAKQEPHFO	479
QY	294	EVEKTAPSGKAREIDESLIFYKKMELACVPDALLATOMDRNALPFTYEODLVYKHK	353
Db	480	SVARDIHSSGSLGSEIRDRDTSCTGCSARKOTADVTIDVARMKEL-----FSETSLLKHK	533
QY	354	LDLELYPGGPESVYIQLHGYFLFKMSPEDIRKMNVTSLETK-----ALLEVDKG	402
Db	534	VGE--PATRTINVPD-----AOSPGEH--NLHVTETKAEEDSGLKNQOALYNLSKA	582
QY	403	HEMSPQAPRRRLPQVAL-----IDRF	424
Db	583	DKLVS DIP-HPVEDLTTSGSVANKNDVIGSSKVAENELVYKIPGGDVDSYIQLSGNT	641
QY	425	VKRGQDQDKDTIDL-----TAFYPGYCLSLPEELSSV-----PPSSIAVNR	467

```

Db      642 LTAASLEKCTADQLGKLSQEGETTPASDGETCHLAEEASLSLYRSEPTASA----- 697
Oy      468 PODLDTCDDPRLDVLTKKARLAFON-----MNGSEYFKIQSLGALPTEDLKALSOONVS 523
Db      698 ---STAEPLPTDKL--EKNISFODEVKTLLNGD---KREAIL-----LSSEQOTVWN 741
Oy      524 MDLAT---FMKLRDVAVLPLTVAEYOKLGLGPHVEGLKAEERHVRPD 567
Db      742 SKITNSEELQASTDEV-----PHTDGKSVDAVNAQTVKE 776

```

## RESULT 4

```

T03455
ALR protein - human
C:Species: Homo sapiens (man)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03455
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A:Reference number: Z14954; MUID:97388474
A:Accession: T03455
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4957 <PRA>
A:Cross-references: EMBL:AF010404; NID:g2358286; PID:ACG51735.1; PID:g2358287
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: human ALR protein
C:Keywords: alternative splicing

```

```

Query Match      3.48; Score 111.5; DB 2; Length 4957;
Best Local Similarity 19.9%; Pred. No. 74;
Matches 148; Conservative 112; Mismatches 287; Indels 195; Gaps 34;

```

```

Oy      8 PCWCGDRPGSLFLFLSLGWVNP--ARTLAGETGESAPLGVLTPPHNISLSRQLL 65
Db      3557 PSVSLGDQPGSMTONL--LGFOQRMLEKPMONNTGPPRKGPVLQSGGLPGVGINPTV 3614
Oy      66 GFPCAENVSG-----LSTERYRELAVALAOKNVKLTSTOILKSLAHRLSEPPD--L 113
Db      3615 GOLRAQOLGVLAKNPOLNHLSPQOOQOLQALLMORLOQSO-----AVROTTPYQBERGT 3668
Oy      114 DALPLDLLFLNPD--AFSGQACTRFSSRTTKANVDLLPRGAREKQYLLAALACWYR 171
Db      3669 QTSPLQGLGCGPOLGGRPGPT---GPIQELGAGRRPG--PRRLPAPGALSTGPLY 3722
Oy      172 GSL-----LSEADVRAALGGLCDLPGRFVAESAELL-----PRLVSC 209
Db      3723 GVINPTPPSSPQPERK-----PSQLPSPSSQLPTREADLPRTNHTGTRKPGQETLEPP 3774
Oy      210 PGLPDQDOEARAALOGGRRPYRPTWSVTMDALRGLR-----VLGQPTIRSIPOG 264
Db      3775 PGRVSPAQAOLADTLFSGKLPMDPRDMLAETQKPEQSSLVPGHLDVNGO-----VVEEA 3830
Oy      265 IVAAMRQSSSDPMSRQERTILRRFRREVE-----KTACPSGKKAREIDE 311
Db      3831 SOLSTKQPREPC-----ALGAOSVREANGERTIGARGTSMHLLLAGPSEAGHLLQ 3884
Oy      312 SLIFPKKMLACVADALLATQMD--RVNAIPFTYEOLDVLLKHKIDELTPQGYRESVIOHL 370
Db      3885 KILRAKNVQLSTGGSGSEDLRAEINCHIDS-----KLAGLEQKL-----QGTSPSKEDAA 3933
Oy      371 GYLFLKMSPEDIRKWN---VTSLETLK-----ALLEVDKGHEMSPPQARRPLPOVA 418
Db      3934 ARKLPDTPKPKHVOKASDRLVSSRKLRKEDGVRASEALLK-----QLKQELSLPLTGPBA 3988
Oy      419 -----TLIDRFVKG-----RGQDKDVLDTLTATFYPEYL-----CSLSEELSSVPP 460
Db      3989 ITANFSLPAPGSGCPVNGQSOLRGAFSGGALPTGPDYYSOLLTKNNLSNPPTPPSSLP 4048

```

```

Oy      461 SSIMAVRPODDTCDDPRLDVLTKKARLAFON-----MNGSEYFKIQSLGALP----- 510
Db      4049 TPPSVQOKMKNVGTSPSEELGEHPRKDAASADSERALROTSE--VKSIDLLALPTPHN 4106
Oy      511 -TEDLKALSOONVSMDLATEFMKLRDVAVLPLTVAEYOKLGLG-----PHVGLKAEERH 563
Db      4107 QTEDEVRESDESD-----SPDSIVPASPE--SIIGEAPRFRPHLAGSGRMQEDR 4155
Oy      564 FVRDMLRQRODDL-----DT-----LGIGLOGGIP-----NGVLYLDLSOET 602
Db      4156 ALSPIVPLIPRDSIPVPPDTPYGALGLEVPGLPYTTEKKGSGSEVSVMLTVAADKN 4215
Oy      603 LSGTRCLLGPVLTVALALLA 624
Db      4216 LNG-----VMVAVALLS 4228

```

## RESULT 5

```

T03454
ALR protein - human
C:Species: Homo sapiens (man)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03454
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Ya
Oncogene 15, 549-560, 1997
A:Title: Structure and expression pattern of human ALR, a novel gene with strong homo
A:Reference number: Z14954; MUID:97388474
A:Accession: T03454
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5262 <PRA>
A:Cross-references: EMBL:AF010403; NID:g2358284; PID:ACG51734.1; PID:g2358285
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: human ALR protein
C:Keywords: alternative splicing

```

```

Query Match      3.48; Score 111.5; DB 2; Length 5262;
Best Local Similarity 19.9%; Pred. No. 81;
Matches 148; Conservative 112; Mismatches 287; Indels 195; Gaps 34;

```

```

Oy      8 PCWCGDRPGSLFLFLSLGWVNP--ARTLAGETGESAPLGVLTPPHNISLSRQLL 65
Db      3662 PSVSLGDQPGSMTONL--LGFOQRMLEKPMONNTGPPRKGPVLQSGGLPGVGINPTV 3919
Oy      66 GFPCAENVSG-----LSTERYRELAVALAOKNVKLTSTOILKSLAHRLSEPPD--L 113
Db      3920 GOLRAQOLGVLAKNPOLNHLSPQOOQOLQALLMORLOQSO-----AVROTTPYQBERGT 3973
Oy      114 DALPLDLLFLNPD--AFSGQACTRFSSRTTKANVDLLPRGAREKQYLLAALACWYR 171
Db      3974 QTSPLQGLGCGPOLGGRPGPT---GPIQELGAGRRPG--PRRLPAPGALSTGPLY 4027
Oy      172 GSL-----LSEADVRAALGGLCDLPGRFVAESAELL-----PRLVSC 209
Db      4028 GVINPTPPSSPQPERK-----PSQLPSPSSQLPTREADLPRTNHTGTRKPGQETLEPP 4079
Oy      210 PGLPDQDOEARAALOGGRRPYRPTWSVTMDALRGLR-----VLGQPTIRSIPOG 264
Db      4080 PGRVSPAQAOLADTLFSGKLPMDPRDMLAETQKPEQSSLVPGHLDVNGO-----VVEEA 4135
Oy      265 IVAAMRQSSSDPMSRQERTILRRFRREVE-----KTACPSGKKAREIDE 311
Db      4136 SOLSTKQPREPC-----ALGAOSVREANGERTIGARGTSMHLLLAGPSEAGHLLQ 4189
Oy      312 SLIFPKKMLACVADALLATQMD--RVNAIPFTYEOLDVLLKHKIDELTPQGYRESVIOHL 370
Db      4190 KILRAKNVQLSTGGSGSEDLRAEINCHIDS-----KLAGLEQKL-----QGTSPSKEDAA 4238
Oy      371 GYLFLKMSPEDIRKWN---VTSLETLK-----ALLEVDKGHEMSPPQARRPLPOVA 418

```

Db 4239 ARKPLTPKPKRYOKASDBLVSSRRKLRKEDGVASEALLK-----OLKOEISLPLTEPA 4293  
QY 419 -----TLDRFVKG-----RGQLDKDITDITLAFYPGTL-----CSLSEELSSVVP 460  
Db 4294 ITANPSLAPESGCGPVNGQSOLRGAFSGALPTGPDYYSQLLTNNLSNPTPPSSLP 4353  
QY 461 SSIMAVRPQDDTCDPROLDVLYPARLAFON-----MNGSEYFKIQSFLGAP----- 510  
Db 4354 TPPEVQOKMNGVTPSEELGHPKDAASARSEKALNDTSE--VKSIDLALALPTPHN 4411  
QY 511 -TEDIKALSOQNVSMDLATFMKLRDVAVLPTVAEVQKLG-----PHEGLKAEERRH 563  
Db 4412 QTEDEVMESEDSD-----SPDSIVPASPPE--SILGEAPRFPHLGSGRWEQEDR 4460  
QY 564 PYRDLILNRQDDL-----DT-----LGGLGLOGIP-----NGYLVLDLSVOET 602  
Db 4461 ALSVPIPLIPRDSIVFPDTPRYGALGLEVPGLPTTWKKGSEVSMLVSAADKN 4520  
QY 603 LSGFPCLLGPVLTVALLLA 624  
Db 4521 LMG-----VMYAVAEELS 4533

## RESULT 6

hypothetical protein F8A5.20 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: B96632  
R:Title: A.; Becker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huitzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Matzidali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: B96632  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-340 <STO>  
A:Cross-references: GB:AE005173; NID:q2462761; PIDN:AB71980.1; GSPDB:GM00141  
C:Genetics:  
A:Gene: F8A5.20  
A:Map position: 1  
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 3.4%; Score 110.5; DB 2; Length 340;  
Best Local Similarity 20.4%; Pred. No. 2;  
Matches 89; Conservative 62; Mismatches 126; Indels 159; Gaps 22;  
QY 87 LAOKNVKLSLEOLRCLA--HRLSEPPEDLDLPLDLLFLNPDASFSGOACTREFSRITK 144  
Db 3 LGSGLGVASAGLIGCMALSAHYGAKPPTDAIAL-----LHNAIMSG-----VTFP----- 48  
QY 145 ANVDLLPGADERORLLPALALACWGR-----GSLSEADYRALGLACDLPGRFV 195  
Db 49 --PTSDMYGETNELLGKALKDGVKKEVELATKFGFVIEGLISEVRG-----DPEYV 100  
QY 196 AESAELLRL-VSGPG-----PLDQDQEAARALOGG-----GPPYGPST-- 237  
Db 101 RAACEASLRDLACIDLYQHRIIDTRVPIETMELKLVKEEGIKYIGLSEASASTIR 160  
QY 238 -----WSVTMDALRGILPV-----IGOPITRSIP--OGIYAANRO--RSSRD 276  
Db 161 RAHAVHPTTAQVIEKSLMSRAEEDITPICRELIGIGIYAVSLRGFLAAGPKLAENLEN 220  
QY 277 PSHQPERTILRRPREVEKTACPSGKAREIDSLFYRK---WELACVDAALLAT 332

Db 221 DFRKATITLL--PRQOE-----NVDRNKILFEKVSAMARKGCTPQALALA 265  
QY 333 -----QMDRYNAIPFTEQDLVAKHKLDELTPGCEYSYQHLGTYFLKMSPEDIKKMYT 388  
Db 266 VVNHGGDDVCPITPGTK-----TENNNONRALSVKLTPEEI-----S 303  
QY 369 SLETKKALLEVDKGHMSPOAPRRPLPVAVLIDRFVKGRLDKDITDITLAFYPGYIC 448  
Db 304 ELDSL-----APES-----VAGE-----RIMA 321  
QY 449 SLSEELSSVPPSSIW 464  
Db 322 SMSTFKNSNTPLPSM 337

## RESULT 7

T18551  
saframycin Mx1 synthetase B - Myxococcus xanthus  
C:Species: Myxococcus xanthus  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Nov-2000  
C:Accession: T18551  
R:Pospiech, A.; Bietenhader, J.; Schupp, T.  
Microbiology 142, 741-746, 1996  
A:Title: Two multifunctional peptide synthetases and an O-methyltransferase are invol  
A:Reference number: Z18967; MUID:97090395  
A:Accession: T18551  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1770 <POS>  
A:Cross-references: EMBL:U24657; NID:g1171127; PID:g1171128; PIDN:AMC44128.1  
A:Experimental source: strain DM504-15  
C:Genetics:  
A:Gene: safB  
C:Superfamily: acetate--CoA ligase homology; acyl carrier protein homology  
C:Keywords: carrier protein  
F:72-577/Domain: acetate--CoA ligase homology <ACL>  
F:618-668/Domain: acyl carrier protein homology <ACP1>  
F:1222-1659/Domain: acetate--CoA ligase homology <ACL2>  
F:1678-1746/Domain: acyl carrier protein homology <ACP2>

Query Match 3.4%; Score 110; DB 2; Length 1770;  
Best Local Similarity 23.3%; Pred. No. 22;  
Matches 111; Conservative 49; Mismatches 175; Indels 142; Gaps 22;  
QY 1 MALQRLDPCWCGDGRGSLFLFLFSIGVHPR-----FLAGET---GIES 43  
Db 520 VVOGEVDRRYRPGDWPDIYAIIRDISFOHARVHAVVLIKSGILTKTSSGKVGORATRE 579  
QY 44 APLGVLFT-----PHNISSLSPRQLGFPCE-----VSGISTEVRV--ELAVALA 88  
Db 560 AYLBQDLTVSADAQGEVGEFSLSRASLIERSEKRLPALVAYIASATAGSIPRAIL 639  
QY 89 QKNVKLSTEOLRCLA--HRLSEPPEDLD-ALPLDLLFLNPDASFSGOACTREFSRIR 142  
Db 640 TGEVRLPLGLDLSLALVELKRIE---ODLEVALLELRLL-----EGPTLGAVALHL 689  
QY 143 TKANVDLLPRGAP-----ERQRLPALACWGVKSLSEADYRALGLACDLGGRFAVE 197  
Db 690 EAAGREARPVASPTPTGERLVPVPGORALW-----FLHJLADPSGTYTA- 735  
QY 198 SAEVLLPRLVSCPGPLDDQGEAARALOGGPGPYGPSTWSVTMDALRGILPVIGOP 257  
Db 736 -----RAICLOGALD--VEALRRAFOS-----LVTRHPLRATFTFPVAGEP 774  
QY 258 IRSIPQGIYAAMR--ORSSRDPSSWROPERTILRRPREVEKTACPSGKAREIDSLFIY 316  
Db 775 VQGVYAHVPASFLIVDASG-----DEALARRLLRRA-----Y 808  
QY 317 KKWEL-----EACVDALLATQMDRVNVAIPTTYEQDLVAKHKLDELTPGCTPE 364  
Db 809 RPFIDGRGPLRLRTHLFSRAOEHVLLLS---VHHIVDFWSLAVLVDLRLRYEAGGDG 864

QY 365 SVIOHGLYFLKMSPEDIRKWNVTSLTLKALLEVDKGH---EMSPQARRRPLPOVA 418  
 DB 865 VVLPVPG-----QPADAIRWQAEYMAQPSH--BEDMGYMQRLAGPLRLLELPAPA 913

## RESULT 8

T35548  
 hypothetical protein SC6G3\_04 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
 C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
 C:Accession: T35548  
 R:Seeger, K.J.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, July 1999  
 A:Reference number: 221582  
 A:Accession: T35548  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 1-1171 <SEE>  
 A:Cross-references: EMBL:AL096825; PIDN:CAB46976.1; GSPDB:GN00070; SC6G3\_04  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SC6G3\_04

Query Match 3.3%; Score 109; DB 2; Length 1171;  
 Best Local Similarity 20.7%; Pred. No. 14;  
 Matches 163; Conservative 69; Mismatches 246; Indels 310; Gaps 35;

QY 3 LQRL-----DPCWCGDRPGSLFLFLSLGW--VHPARTAGETGESAPLG-CVL 50  
 DB 402 LQRLKCGVGAPEVEVYAGEGGSALTTRMRAAMTPVAARLDVGVGVYAAQAADGTL 461  
 QY 51 TTPH-----NISLSRQLGFP--CAEV-----SGLS-TERVELVALAQ----- 89  
 DB 462 RENHRAAEAGRVTSARVALLGAAARCAITDLHDGITEAREVYPGAALPELLALDL 521  
 QY 90 -----KNVKLSTEOELRLAHL-----SEPPEDLDA----- 115  
 DB 522 LDSVRRRLPPTTEVTRIRARLAGLLDAVRLPGLAGSDERPDAAVYTLAVRSAAD 581  
 QY 116 ---LPLD-----LILFLN-----PDASEGPOACTRF 138  
 DB 582 RLGLRLDELALYSRSGPLQGAQAARVLLDLDGSDLLGRLAGWDVDTAGPGRRL 641  
 QY 139 FSRITKANV-----DLPRGAPRQ 158  
 DB 642 ERRLGVLAAGFLIESASTALGLFEVETMSDRGLDRLYALRGFRALTPQG--RS 698  
 QY 159 RLPLPALCMGVRSGLSEADVRLAGLACDLPGRFVAESAENVLLPRLVSCGPLD--QD 216  
 DB 699 RVLAAVSDRLDPRDLRLPAPALVGRMAA--ADGEGLLALDELGLADLASAPAKADACAT 757  
 QY 217 QOEARAALOG-----GPPYPGPTWSVSTMDALRGLLPVLCQPIIRSIIP--OG 264  
 DB 758 SAPAPRTAPSGAEVPHGAGAPRLGPADRMRLILGHDTAG--LPALRLPYARALDELDFREG 816  
 QY 265 IVAMROSSND-----PS---WQGEERTILRRRF 291  
 DB 817 EBADEESETEGGDDGPEPGAGTSEGSDDRTGGAARSPVSRMAEDLRTLFGAEI 876  
 QY 292 RREVEKTPCGPKARKEIDESLIFYKKWELACVDAALLATQMDRVNAIPPTYEQDLVIA 351  
 DB 877 ROEYIERVAAGRT-----DVIALDPAVSRSVELLSAV----- 911  
 QY 352 HKLDELVPQGPESVYIQLGYFLKMSPEDIRKWNVTSL- TLKALLEVDKCHMSPOAP 410  
 DB 912 ---LTLARGMPQEVASLRLP-VKRLVEELTKELATRLRLPTLTGL-----TTPRPT 958  
 QY 411 RRP-----LPOV--ATL--IDRFVKGROGL--DKDTLDTLAFYGYLCSLPELSVP 459  
 DB 959 RRPGLPLDPLRLRLANLAHIRREDGAVEVPERPVFTRTARRNDMLILIVDVASME 1018

QY 460 PSSIAVVRPDLDTCDPQDLVLYPKARLAEQNMNGSEYFVKIOSFLGAPT----- 511  
 DB 1019 TSVVMS-----ALFAALIGCAPLTSLHFLT 1044

QY 512 ---EDLKALSOQNVSM-----DLATPKLRT--DAVLPITYAEVQKLLGP 551  
 DB 1045 STQVADLTGLVADPLSLLEVEKVGSGTHIAAGLAHARSLVTPPRTLVVVSDFEE--GA 1102

QY 552 HVEGKAE 559  
 DB 1103 AVEGLLAE 1110

## RESULT 9

G82962  
 hypothetical protein PA5464 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: G82962  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laribig, K.; L  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: G82962  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-652 <STO>  
 A:Cross-references: GB:AE004959; GB:AE004091; NID:99951791; PIDN:AGC08849.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA5464

Query Match 3.3%; Score 108.5; DB 2; Length 652;  
 Best Local Similarity 22.6%; Pred. No. 6.8;  
 Matches 98; Conservative 44; Mismatches 143; Indels 149; Gaps 19;

QY 22 LFLSLGWHPARTLAGETGESAPLGVLTPPHNLSLSPQOLGF----- 67  
 DB 177 LLDLAGW--PORT--GQKGRBAHNGFVLRSPDYELSPREFAVNMEYFLLDPSYACR 232  
 QY 68 -----PCAE-----VSGLSTERYVELAVALAOKNVK 93  
 DB 233 RPAITYRYQGRFGAPRHSACAQSFAYLNAGRDGQPLGLQDLPERVYEVYLLAEANDN 292  
 QY 94 LSTE---QLR---CLAHRLSEPPEDLDALPLDLLFLNPDAFSGPOACTRFPSRITKAN 146  
 DB 293 LVSRMGHTMLRLVICAPGRPRGPDRLD--LDQHLVLSYRAFGV----- 334  
 QY 147 VDLPRGAPENORLLPALACW--GYRGSLSSEADVRLAGLACDLPGRFVAESAENVLLPR 205  
 DB 335 -DL-----OLSSMDGLTGAVPSRLFLPLP-----SQVIEEYKVELRS 371  
 QY 206 LVSCGPLDQDQ-----OEARAALOGCGPPYGPSTVSSTMDALRLRLPVLCOPIIRS 260  
 DB 372 LASVPKIDREEVASVYRAAOSHSHSYDGYFTISNNCAVETIKLRSGIFRRRLQSIDSD 431  
 QY 261 I-POGIYAMRORSSRDSMRQPERTILRPFRRE-----VEKTACPSGK-- 304  
 DB 432 ITPGVLEMLENRKLADBSVLDDEPKALRLGYRFDSFDRYOAMFDVLKORHLRPODKVE 491  
 QY 305 ---KAREIDESLIFYKKWELACVDAALLATQMDRVNAIPPTYEQDLVKKHIDEIYP 359  
 DB 492 DWLALPAREROP--WFAKADLRA--SAALLLDQASLR-----ROLLLAODELKLRYL 539  
 QY 360 QGYPESVIOHGLYL 373  
 DB 540 -----LGHLL 543

## RESULT 10





```

QY 154 A-----PRQKILRPAALACMGVRSGLSEADYRALGCIACLPGRFVAESNEVLLPRIVS 208
Db 1275 LDDLSDDERMKIM-----AYMAEADMQNY-----RKPIRGPPMPPSTSM 1315
QY 209 CP-----GLDODDO-----EAAARAALQGGGPYGPSTMSVTAALGLLPVL 253
Db 1316 IPPGMEGLSEEBEROKIMSVMAEATDSSSVITTSKOPSRSP---SVARKOP-QOLMP-- 1367
QY 254 GQPIRSIPOGI-----VAAMRQSSRDSPMSRQPERY----- 285
Db 1368 PQQAIPRIIPGLEGLSDEBRTIMSVMAEAFEEESRQVSP-RQPSRSPSVNPOQSHP 1426
QY 286 -----ILRPFR-REVEKTACPSGCKAREIDSLITKKMEELACYDAALLATQMDRV 337
Db 1427 IPSEPIRPPGLELSEEBEROKIMSVMAEVEESRQSLPQRSRSPSVAMI-QAPAV 1484
QY 338 NAIPPTYQOL-DVLKHKIDELX-----PQCY---PESVIOHGLYLFLKMSPE 380
Db 1485 PIIPSGMEDLEPPAEEROKIMSVMAEAEIONFVPSRSPSVMSQAPVPVIPH-----GLE 1536
QY 381 DIRKANVTSLETKALLEVD-----KGHEMSQAPRRPLPOVATLIDRFYKRGOL 431
Db 1537 DLSEBEROKIISVMAEABIDSAKIIPSRSTSSYMPRLPQMSQPEITTTGLEHSEADMEF 1536
QY 432 DKDLDLTLTAFYPGYLGLSPEELSSVPPSSIWAIVRPODLDTCPR-OLD-VLYPKARL 488
Db 1597 GRDSRSRQVPIPG-LENLSEEBERQI--MSYMA--HALESIIPSGHIDPISLPRHT 1651
QY 489 AFQÑ---MNGSEBYFYKIOSFLAGAPTEDLKALISQONVSMDLATEFKLTDAVLPITVAEV 545
Db 1652 GFKPAGIYNEDLEFETERKOREESPTRRESGYATSTSYERELAMGDEEBRMDGLE-DITRI 1710
QY 546 QKLLGPHVEGLKAEBHRPVRDWILRQRODD 576
Db 1711 R-----EGARS---RDRSDREYVLRREED 1731

```

RESULT 13  
E82652 ABC transporter ATP-binding protein Xf1602 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: E82652  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: E82652  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-283 <SIM>  
A:Cross-references: GB:AE003987; GB:AE003849; NID:99106634; PIDN:AAF84411.1; GSPDB:GN001  
A:Experimental source: Strain 9a5c  
R:Stimpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Brigues, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carner, H  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junguelza, M.L.; Kemper, E.L.; Kitzajma, J.P.; Krieger, J.E.; Kurame, E.E.; Laigis  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, J.M.F.; Martino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.Y.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sasanaka  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
A:Genetics:  
A:Gene: Xf1602

[illegible]

RESULT 14

T42756

5-oxoprolinase (ATP-hydrolyzing) (EC 3.5.2.9) - rat

N.Alternate names: pyroglutamate (ATP-hydrolyzing)

C.Species: Rattus norvegicus (Norway rat)

C.Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-May-2000

C.Accession: T42756

R.Ye, G.J.; Breslow, E.; Meister, A.

J. Biol. Chem. 271, 32293-32300, 1996

A.Title: The amino acid sequence of rat kidney 5-oxo-L-prolinase determined by cDNA

A.Reference number: 222263; MUID:97113037

A.Accession: T42756

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: mRNA

A.Residues: 1-1288 <YEG>

A.Cross-references: EMBL:U70825; NID:g1732064; PID:g1732065; PIDN:AAC52955.1

A.Experimental source: strain Sprague-Dawley; kidney

C:superfamily: hypodietical protein YKL215c

C:keywords: homodimer; hydrolase

Query Match	3.3%	Score 107.5	DB 2	length 1288	
Best Local Similarity	20.4%	Pred No. 21			
Matches 140	Conservative 69	Mismatches 224	Indels 253	Gaps 34	
Qy	19	LLFLFLSLGWHPARTLAGETGESAPLGGVL--TTPHNISLSRLQLGFPACAEVSL	75		
Db	265	YIFMR-SDGGLPMPAFSGSRALVSGPAGVGVSATYHNLGGCP--YIGF--DMGCT	318		
Qy	76	ST-----RRVELNALAKNNKLSLEOURLCLANHLSEPPEDLDLPL	118		
Db	319	STDVSRVAGEEFHVEEASTA---GVTLDAFOL-----DINVAAGGSRLEF	362		
Qy	119	-DLLFLFND--AFSGPACTRFPSRITKANVDLLPGAPEROLLPALACMGVRSGL	174		
Db	363	RSGLFVVGESAGAHGCP-ACYRKGGPVTVDANLV-----LGRLLPASFPC-----	408		
Qy	175	LSEADVRLGLACDPLGRFVAESAELLFLVSCGPPLDODQOEARARALOGCGPPY--	232		
Db	409	-----ITPGEQGPLSPASRKALAEVVAEIVNS	436		
Qy	233	---GPPTSWSYTMDALRGLLPLVLCQPIIRSIPOGIYAAMRORSSRDPSPROPERTILR	288		
Db	437	FLTNPCPASQSLSELEVAMGFVRVNAEMACRPI---RALQARQHDSS-----	481		

QY 289 PRFRREVETACPSG-----KKAREIDSLIFYKKMELE---ACVDAALLATQMD 335  
 Db 482 -----AHVLAFCGAGQGHACALARALGMDTVHHRHSGLLSALGLADLVYHNAQER 535  
 QY 336 RVNAIPETEDLDVKKHKLDE---LYPGQYPSVIGHLYFLKMSPEDIRKWNVTSL 390  
 Db 536 SLSTYPEFADLDORLSLREQCVDALQVGFPRSOISTESFHLHRYGTDC----- 587  
 QY 391 ETIKALLEVDKHEMSPPARRP-----LPQVATLIDRF-VKRG--- 429  
 Db 588 -----ALWMSAHQHDPATACSPRAGDGAFAVERMYREFFGFIIPRPVVDDVVRGTGRSG 643  
 QY 430 -QLDKDT-----LDTLTAFY-----PGYL----- 447  
 Db 644 LQLE-DTFKIQGTGPHVKEVYKQCFEGEGYQETFPVYLLGELGNGHOLQPCILLIDNNSTIL 702  
 QY 448 -----CSLSPBELSSVPPSSIAVPRQDLDLT-CDPRQDLVLYPKARLAFONNNGSEYFVKI 502  
 Db 703 VEPGQAEVETDTGDI-RISVGAEGSPSMADTRIDPIQLSIF-----SHREFSI 748  
 QY 503 OSFLGAPTEDELKALSOQNVSMDLATFPMKLRDPAVLPLTVAEYOKLLGPHVEGLKAEBRH 562  
 Db 749 AEQMG-----RIIQRAIS-----TNKERLDPSG-----ALFGPD-GGLVSNAPH 788  
 QY 563 RPRVDMLIRQR-QDDLDLTGLGLOGG 587  
 Db 789 IPVHLGAMQETVQFOIHLGADLHPG 814

## RESULT 15

S23070

erythronolide synthase (EC 2.3.1.94) II - Saccharopolyspora erythraea

N:Alternate names: 6-deoxyerythronolide B synthase II

C:Species: Saccharopolyspora erythraea

C:Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text-change 21-Jul-2000

R:Accession: S23070; S22011; S23205

R:Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.

Eur. J. Biochem. 204, 39-49, 1992

A:Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of

A:Reference number: S23070; MUID:92155230

A:Accession: S23070

A:Molecule type: DNA

A:Residues: 1-5573 &lt;BEV1&gt;

A:Cross-references: EMBL:X62569

A:Experimental source: strain NRRL 2338

R:Bevitt, D.J.

submitted to the EMBL Data Library, September 1991

A:Reference number: S22011

A:Accession: S22011

A:Molecule type: DNA

A:Residues: 1-184,'I',186-301,'S',303-521,523-658,'A',660-993,1001-1212,'H',1214-1392,13

3479,'DH',3480-3572 &lt;BEV2&gt;

A:Cross-references: EMBL:X62569; NID:946977; PIDN:CA44448.1; PID:9581651

R:Caiffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.

FEBS Lett. 304, 225-228, 1992

A:Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of th

A:Reference number: S23103; MUID:92316235

A:Accession: S23205

A:Molecule type: protein

A:Residues: 2-12,'XXX' &lt;CAP&gt;

A:Experimental source: strain CA340

C:Genetics:

A:Gene: eryA

A:Start codon: GTG

C:Function:

A:Description: catalyzes the construction of a polyketide chain, which is then cyclised

A:Pathway: erythromycin biosynthesis

C:Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car

ort-chain alcohol dehydrogenase homology

C:Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; multi

F:52-453/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology &lt;OAS1&gt;

F:561-843/Domain: [acyl-carrier-protein] S-malonyltransferase homology &lt;AMT1&gt;

F:1140-1308/Domain: short-chain alcohol dehydrogenase homology <SAD2>  
 F:1404-1475/Domain: acyl carrier protein homology <ACP>  
 F:1519-1919/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>  
 F:2023-2305/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>  
 F:2857-3131/Domain: long-chain alcohol dehydrogenase homology <LADH>  
 F:3149-3327/Domain: short-chain alcohol dehydrogenase homology <SADH>  
 F:3420-3493/Domain: acyl carrier protein homology <ACP>  
 F:1439/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 3.38; Score 107.5; DB 2; Length 3573;

Best Local Similarity 20.48; Pred. No. 90;

Matches 147; Conservative 63; Mismatches 225; Indels 285; Gaps 35;

QY 10 WSCGRPSLFLFLFSLGVHPARTLAGETGTSAPLGGVLTTHNTSSLSRQLLGPC 69  
 Db 1894 WPAGERP-----RRACVSEFGVSG--TNAHYVEAP----- 1923  
 QY 70 AEVSGLSERVRELAVALAOKNVKLSFEOLCLAHRLSEPPEDDALPLDLLFLNPDAF 129  
 Db 1924 AEQEAFTER-GPIPEFVLSGRSEAVVAQAARALAEHLDTPE-----LGUTDAW 1972  
 QY 130 SGPOACTFFFSRITK-----ANVDLPRGAPERKRLPALACGVGSLSEADY 180  
 Db 1973 TLARGARFVDRAAYLVGDRAVCALDIALAEGRSADAVAPVTS-----PKRPV 2023  
 QY 181 RALGLACDLPG--RFVESAELLRLVSCPPLPDQQAARALQGGCPGPSTW 238  
 Db 2024 LVFPGQAGQWGMARDLLESSEVFESMSRCAPL-----SPHTDW 2064  
 QY 239 SVSTMALRG-----LPLVGOPITRISIPQIVAAARORSSRDSWROPERTILRR 290  
 Db 2065 KL--LDVVRGGGDPDPRHVRVVL-OPVLFSTIMVSLAEIMRAHG-----VTP- 2107  
 QY 291 FRREVETACPSGKKARIDSLIFYKKMELEACVDALLTOMDRVNAIPFTEQDLVL 350  
 Db 2108 -----AANVGHSGGTAAAHV-AGALSTENAAKVVALNSQV----- 2142  
 QY 351 KHKIDELYPQYPSVIGHLYFLKMSPEDIRKWNVTSLETLKA----- 395  
 Db 2143 --LRELDQG-----GMVSVGSARDE-----LETVLARMDGRVAVAVNGPG 2182  
 QY 396 -----LLEVD-----KGHEMSPO--APR--RPLQVATLIDRFVKGRLDKDITDT 438  
 Db 2183 TSVVAGPTAEIDFEFAEAEMKPRRIAYVASHSPVARIEDRLA-----ELGT 2234  
 QY 439 LTAFFPYGLCSLSPBELSSVPPSSIAVPRQDLDLTCDPRQDLVLYPKARLAFONNNGSEY 498  
 Db 2235 ITAV-----KGSVPLHS--TVTGEVDT--SANDAST-----WKNLRPRVL 2272  
 QY 499 FVKIQSLGAPTEDELKALSOQNVSMDLATFPMKLRDPAVLPLTVAE----- 544  
 Db 2273 F-----EQAVRELTVGGFD--TFEVRSPHVLMAVEENAEHGAECVTCPT 2317  
 QY 545 -----VQKLGHVGLKAEBR--HRPV-----RDWLIRORODLD 578  
 Db 2318 LRREGSGPHEFLRLRAHVHGVGADLRPAVAGRPALPTYPPEHGFMRPRPADVS 2377  
 QY 579 TLGL-----GLGGIPNGVYLVDLS--VOETLSCTPCLLGGPLVTVAL 621  
 Db 2378 ALGVRGAEPHLLLAANDVYVGHGAVFTGRSLTDEQPLAEHVVGRTLY-PGSVLVDLAL 2436

Search completed: December 7, 2001, 10:02:37

Job time: 610 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 09:47:57 ; Search time 23.78 Seconds  
(without alignments)  
968.272 Million cell updates/sec

Title: US-09-215-035-2

Perfect score: 3261  
Sequence: 1 MALGRUDPCWCGDRPGSL.....LLGPPVLVALLLASTIA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3261	100.0	628	1 MSLN_HUMAN	Q13421 homo sapien
2	1223	3.8	650	1 APP1_HUMAN	P16163 homo sapien
3	111	3.4	1108	1 CYGE_MOUSE	P32785 mus musculu
4	110.5	3.4	5065	1 EPPL_HUMAN	P8107 homo sapien
5	109.5	3.4	834	1 DPOL_THCEA	P80194 thermus aqu
6	108.5	3.3	1398	1 DNA2_SCHPO	Q9ur2 schizosacch
7	107.5	3.3	476	1 AUPL_HUMAN	Q9y679 homo sapien
8	107.5	3.3	1288	1 ERY2_HUMAN	P97608 ratu
9	107.5	3.3	3567	1 ERY2_SACER	Q03132 saccharopol
10	106.5	3.3	653	1 APPI_MOUSE	Q03157 mus musculu
11	106	3.3	690	1 HRPO_BURSO	P35656 burkholderi
12	105	3.2	693	1 PKNL_MYXXA	P33973 myxococcus
13	104.5	3.2	834	1 DPOL_THETH	P52028 thermus aqu
14	104	3.2	1003	1 SYG_CHLTR	Q46371 chlamydia t
15	104	3.2	1142	1 SPK_HUMAN	Q92797 homo sapien
16	103.5	3.2	4128	1 PRKD_MOUSE	P97313 mus musculu
17	102	3.1	1159	1 DP3A_HAEIN	P43743 haemophilus
18	101	3.1	677	1 LCN_HUMAN	P81274 homo sapien
19	100.5	3.1	1073	1 RAG1_ONCMY	Q91187 oncorhynch
20	99.5	3.1	461	1 NCB1_HUMAN	Q02818 homo sapien
21	99.5	3.1	1807	1 ITB4_RAT	Q64633 ratu
22	98.5	3.0	785	1 SYFB_THETH	P27002 thermus aqu
23	98.5	3.0	847	1 IP3L_HUMAN	P27987 homo sapien
24	98	3.0	917	1 IL6B_MOUSE	Q00560 mus musculu
25	98	3.0	1103	1 CYGD_HUMAN	Q02846 homo sapien
26	98	3.0	1108	1 CYGE_RAT	P51840 ratu
27	97.5	3.0	344	1 VIB1_AGRTS	P07169 agrobacteri
28	97.5	3.0	574	1 MP12_RAT	P48966 ratu
29	97.5	3.0	1762	1 DPOO_HUMAN	Q75417 homo sapien
30	97.5	3.0	2035	1 EVPL_MOUSE	Q96952 mus musculu
31	97	3.0	780	1 SAV_SULAC	Q07590 sulfobolus
32	97	3.0	918	1 IL6B_RAT	P40190 ratu
33	97	3.0	1109	1 CYGD_CANFA	O19179 canis fami

34	96.5	3.0	308	1 GDFF_HUMAN	Q99988 homo sapien
35	96.5	3.0	494	1 GP2A_ECOLI	P25552 escherichia
36	96.5	3.0	984	1 PKL2_HUMAN	O16513 homo sapien
37	96	2.9	367	1 CURD_STRON	Q02587 streptomyce
38	96	2.9	919	1 DNL1_HUMAN	P18858 homo sapien
39	95.5	2.9	513	1 DNL1_YEAST	P32340 saccharomyc
40	95.5	2.9	5035	1 RYNR_PIG	P16960 sus scrofa
41	95	2.9	991	1 BMPL_MOUSE	P98063 mus musculu
42	95	2.9	1159	1 DP3A_PASMU	Q9CPK3 pasteurella
43	95	2.9	1164	1 YAV1_XANCV	P14727 xanthomonas
44	95	2.9	1698	1 Y076_HUMAN	Q14999 homo sapien
45	95	2.9	2241	1 TEGU_HCMVA	P16785 human cytom

## ALIGNMENTS

RESULT	1	MSLN_HUMAN	STANDARD	PRT	628 AA.
AC	Q13421	MSLN_HUMAN			
DT	30-MAY-2000	(Rel. 39, Created)			
DT	30-MAY-2000	(Rel. 39, Last sequence update)			
DT	30-MAY-2000	(Rel. 39, Last annotation update)			
DE	MESOTHELIN PRECURSOR (CAK1 ANTIGEN).				
GN	MSLN.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID:9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Carcinoma;				
RX	MEDLINE=96133892; PubMed=8552591;				
RA	Chang K., Pastan I.;				
RT	"Molecular cloning of mesothelin, a differentiation antigen present				
RT	on mesothelium, mesotheliomas, and ovarian cancers.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 93:136-140(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.				
RX	MEDLINE=94200897; PubMed=8150545;				
RA	Chang K., Pastan I.;				
RT	"Molecular cloning and expression of a cDNA encoding a protein				
RT	detected by the K1 antibody from an ovarian carcinoma (OVCA-3) cell				
RT	line.";				
RL	Int. J. Cancer 57:90-97(1994).				
CC	-I- FUNCTION: MAY PLAY A ROLE IN CELLULAR ADHESION. ANTIGENIC PROTEIN				
CC	REACTIVE WITH ANTIBODY K1.				
CC	-I- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.				
CC	NOT SHED TO THE SERUM OF CANCER PATIENTS.				
CC	-I- TISSUE SPECIFICITY: EXPRESSED IN MESOTHELIAL CELLS, MESOTHELIOMAS,				
CC	OVARIAN CANCERS, AND SOME SQUAMOUS CELL CARCINOMAS.				
CC	-I- PTM: GLYCOSYLATED.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
CC	EMBL: U40434; AAC50348.1; -				
DR	MIM: 601051; -				
DR	MIM: 601051; -				
KW	Signal; Glycoprotein; Antigen; GPI-anchor.				
FT	SIGNAL	1	31	POTENTIAL.	
FT	PROPEP	32	294	REMOVED IN MATURE FORM (POTENTIAL).	
FT	CHAIN	295	586	MESOTHELIN.	
FT	PROPEP	587	628	REMOVED IN MATURE FORM (POTENTIAL).	
FT	CARBOHYD	55	55	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	386	386	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	494	494	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	521	521	N-LINKED (GLCNAC. . .) (POTENTIAL).	

FT LIPID 586 586 GPI-ANCHOR (POTENTIAL).  
SQ SEQUENCE 628 AA: 69044 MM: 2783ED29C3855F5D CRC64:

Query Match 100.0%; Score 3261; DB 1; Length 628;  
Best Local Similarity 100.0%; Pred. No. 1.2e-221;  
Matches 628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MALQLDPCWCGDRPGSLFLFLSLGWHPARTLAGGMPFASPLGVLTPHNISSIS 60  
DB 1 MALQLDPCWCGDRPGSLFLFLSLGWHPARTLAGGMPFASPLGVLTPHNISSIS 60  
OY 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNKVLSLSTEDQLCLAHRLSEPPEDLALPLDL 120  
DB 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNKVLSLSTEDQLCLAHRLSEPPEDLALPLDL 120  
OY 121 LFLNPDAFSGPOACTRFRFSRTTKANVDLLPRGAPERQRLPALALACWVSGSLSEADY 180  
DB 121 LFLNPDAFSGPOACTRFRFSRTTKANVDLLPRGAPERQRLPALALACWVSGSLSEADY 180  
OY 181 RALGGLACDLPGRFVAESAENVLLPRLVSCPGPLDDQQAARALOGGGPPYGPSTMSV 240  
DB 181 RALGGLACDLPGRFVAESAENVLLPRLVSCPGPLDDQQAARALOGGGPPYGPSTMSV 240  
OY 241 STMALRGLPLVGPPIRSIPQGIYAAMRORSRDPSPMRQPERTILPRFRREVEKTAC 300  
DB 241 STMALRGLPLVGPPIRSIPQGIYAAMRORSRDPSPMRQPERTILPRFRREVEKTAC 300  
OY 301 PSGKAREIDSLFYKKWELEACYDAALLATQMDRVNAIPPTYQOLDVYKAKLDELTPQ 360  
DB 301 PSGKAREIDSLFYKKWELEACYDAALLATQMDRVNAIPPTYQOLDVYKAKLDELTPQ 360  
OY 361 GYPSVIGHLGLFLKMSPEDIRKKNVTSLETALKALLEKDKHEMSPOAPRRPLQVATL 420  
DB 361 GYPSVIGHLGLFLKMSPEDIRKKNVTSLETALKALLEKDKHEMSPOAPRRPLQVATL 420  
OY 421 IDRFVKGSGQDKDTLDLTLTAEPGYLCSLSPEELSVSPSSIMAVRQDDLTCDPROLD 480  
DB 421 IDRFVKGSGQDKDTLDLTLTAEPGYLCSLSPEELSVSPSSIMAVRQDDLTCDPROLD 480  
OY 481 VLYPKARLAFQNMNNGSEFVFKIQSLGAPTEDLKALSOQVSMPLAFMKRTAVPL 540  
DB 481 VLYPKARLAFQNMNNGSEFVFKIQSLGAPTEDLKALSOQVSMPLAFMKRTAVPL 540  
OY 541 TVAEVQKLLGPHVEGLKAEERHPRVDMILRQDDDLTLGLGOGGIPNGLYVLDSLQ 600  
DB 541 TVAEVQKLLGPHVEGLKAEERHPRVDMILRQDDDLTLGLGOGGIPNGLYVLDSLQ 600  
OY 601 ETLGTPCLLGGPVLVALLASTLA 628  
DB 601 ETLGTPCLLGGPVLVALLASTLA 628

RESULT 2

APPL\_HUMAN STANDARD; PRT: 650 AA.  
ID APPL\_HUMAN STANDARD; PRT: 650 AA.  
AC P51693; 000113;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE AMYLOID-LIKE PROTEIN 1 PRECURSOR (APPLP).  
GN APPLP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98088960; PubMed=9428684;  
RX Faliga K., Peraus G., Kieger S., Duvrignat U., Hesse L., Multhaup G.,  
RX Masters C.L., Beyreuther K., Weidemann A.,  
RT "Human amyloid precursor-like protein 1-cDNA cloning, ectopic  
expression in COS-7 cells and identification of soluble forms in the

RT cerebrospinal fluid.";  
RL Eur. J. Biochem. 250:354-363(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98180887; PubMed=9521588;  
RX Leinkert U., Kestila M., Lamerdin J., McCready P., Adamson A.,  
RX Olsen A., Trygvason K.;  
RT "Structure of the human amyloid-precursor-like protein gene APPL1 at  
RT 19q13.1".  
RL Hum. Genet. 102:192-196(1998).  
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. C-TERMINALLY  
CC PROCESSED IN THE GOLGI COMPLEX AND IS THEN SECRETED.  
CC -I- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN, PARTICULARLY  
CC IN THE CEREBRAL CORTEX POSTSYNAPTIC DENSITY.  
CC -I- PTM: N- AND O-GLYCOSYLATED.  
CC -I- SIMILARITY: BELONGS TO THE APP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U48437; AAB96331.1; -  
DR EMBL: AD000864; AAB50173.1; -  
DR MIM: 104775; -  
DR InterPro: IPR001868; A4\_APP.  
DR Pfam: PF02177; A4\_EXTRA; 1.  
DR PRINTS: PR00203; AMYLOIDA4.  
DR SMART: SM00006; A4\_EXTRA; 1.  
DR PROSITE: PS00319; A4\_EXTRA; 1.  
DR PROSITE: PS00320; A4\_INTRA; 1.  
KW Glycoprotein; Transmembrane; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 650  
FT DOMAIN 22 580  
FT TRANSMEM 581 603  
FT DOMAIN 604 650  
FT DOMAIN 640 643  
FT DOMAIN 241 247  
FT DOMAIN 264 268  
FT CARBOHYD 337 337  
FT CARBOHYD 461 461  
FT CARBOHYD 551 551  
FT CONFLICT 48 48 P->A (IN REF. 2).  
SQ SEQUENCE 650 AA: 72202 MM: 121A034B708C67CA CRC64:  
Query Match 3.8%; Score 123; DB 1; Length 650;  
Best Local Similarity 19.1%; Pred. No. 0.37;  
Matches 130; Conservative 79; Mismatches 229; Indels 242; Gaps 30;  
OY 100 RCLAHRLSEPPEDLALPLDLFLF-----LNP 126  
DB 9 RGLSRPQGPPLPL-LPLLLLLRAOPAIGSLAGSPGAPAPSAQVAGICGLTLHR 67  
OY 127 DAFSG-----PQA-----CTRFPSRTTKANVDLLPRGAPERQRLPALALAC 167  
DB 68 DLRTRWEPPDQSRRCRLDPQRLVLEYCRQMPPELQIARVE-----QATQALPMEWC 120  
OY 168 WGVRSGLSEADVRALGGLACDLPGRFAESAENVLLP-----RLVSCGPL--DOD 216  
DB 121 GGSRSGLCAHPHNOV-FPRC-LPGEFVSEA--LLVPEGCRFLHBERDQCESSRRHOE 176  
OY 217 QOEAAARA--ALOGGG-----PPYGPSTMSVSTMALRGLPLVLGQ 255  
DB 177 AQEACSSOGLILHSGMMLPCGSDRFRGVEYVCCPPGTPPDSGTA-----VGD 225  
OY 256 PIIRSIPOGIYAAMRORSRDPSPMRQ--ERTILPRFRREVEKTACPSGKAREIDSLI 314  
DB 226 PSTRSWPSPGSRVEGADEDEEESFPQVDYFVEPPQAEDEEETVPPSSHTLAV----- 280



```

QY 437 DTAAFYBGLCSLSPED-----LSSVPP-----SSIAVVRPODIDT 473
DQ 741 EVVCGSTPYAMLELPEVIRVRSPPRLCRPLVSMQDAPMCILMPCNAHEHLRPS 800
QY 474 CD-----PROLDVLYPKARLAFQNNNGSEYFKVIOSEFLGAPLEDIKALISQNV 522
DQ 801 MDLFFDLKRSINKRKRTNIIDSMRLMEIOYSNMLDLIR-----ERTELEJOEKO 850
QY 523 SMDLAFMKLRDVAFLPLTVAEVOKLGPHYEGELKAE 560
DQ 851 KTD-----RLLTQ-MLRPSVAEALK-MGTSVEPEYFEE 861

RESULT 4
EPPL_HUMAN STANDARD; PRT: 5065 AA.
ID EPPL_HUMAN
AC P58107;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE EPIPLAKIN (450 KDA EPIDERMAL ANTIGEN).
GN EPIK1 OR EPIPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=21201183; PubMed=11278896;
RA Fujiwara S., Takeo N., Ohtani Y., Parry D.A.D., Kunimatsu M., Lu R.,
RA Sasaki M., Matsuo N., Khaleduzzaman M., Yoshioaka H.;
RT "Epiplakin, a novel member of the plakin family originally identified
RT as a 450-kDa human epidermal autoantigen: structure and tissue
RT localization."
RL J. Biol. Chem. 276:13340-13347(2001).
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN LIVER,
CC SMALL INTESTINE, COLON, SALIVARY GLANDS, STOMACH AND APPENDIX.
CC -1- SIMILARITY: CONTAINS 65 PLECTIN REPEATS.
CC -1- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB051895; BAB40803.1;
KW Coiled coil; Repeat; Structural protein; Cytoskeleton.
CC
FT REPEAT 9 46 PLECTIN 1.
FT REPEAT 47 84 PLECTIN 2.
FT REPEAT 85 122 PLECTIN 3.
FT REPEAT 123 160 PLECTIN 4.
FT REPEAT 166 200 PLECTIN 5.
FT REPEAT 253 290 PLECTIN 6.
FT REPEAT 291 328 PLECTIN 7.
FT REPEAT 330 366 PLECTIN 8.
FT REPEAT 367 404 PLECTIN 9.
FT REPEAT 521 558 PLECTIN 10.
FT REPEAT 578 615 PLECTIN 11.
FT REPEAT 616 653 PLECTIN 12.
FT REPEAT 654 691 PLECTIN 13.
FT REPEAT 692 729 PLECTIN 14.
FT REPEAT 733 767 PLECTIN 15.
FT REPEAT 840 878 PLECTIN 16.
FT REPEAT 898 935 PLECTIN 17.
FT REPEAT 936 973 PLECTIN 18.
FT REPEAT 974 1011 PLECTIN 19.
FT REPEAT 1012 1049 PLECTIN 20.
FT REPEAT 1214 1251 PLECTIN 21.
FT REPEAT 1252 1289 PLECTIN 22.

```

```

FT REPEAT 1290 1327 PLECTIN 23.
FT REPEAT 1328 1365 PLECTIN 24.
FT REPEAT 1366 1403 PLECTIN 25.
FT REPEAT 1539 1576 PLECTIN 26.
FT REPEAT 1577 1614 PLECTIN 27.
FT REPEAT 1615 1652 PLECTIN 28.
FT REPEAT 1653 1690 PLECTIN 29.
FT REPEAT 1694 1728 PLECTIN 30.
FT REPEAT 1865 1902 PLECTIN 31.
FT REPEAT 1903 1940 PLECTIN 32.
FT REPEAT 1941 1978 PLECTIN 33.
FT REPEAT 1979 2016 PLECTIN 34.
FT REPEAT 2017 2054 PLECTIN 35.
FT REPEAT 2192 2234 PLECTIN 36.
FT REPEAT 2235 2272 PLECTIN 37.
FT REPEAT 2273 2310 PLECTIN 38.
FT REPEAT 2311 2348 PLECTIN 39.
FT REPEAT 2352 2386 PLECTIN 40.
FT REPEAT 2726 2768 PLECTIN 41.
FT REPEAT 2769 2806 PLECTIN 42.
FT REPEAT 2807 2844 PLECTIN 43.
FT REPEAT 2845 2882 PLECTIN 44.
FT REPEAT 2886 2920 PLECTIN 45.
FT REPEAT 3260 3302 PLECTIN 46.
FT REPEAT 3303 3340 PLECTIN 47.
FT REPEAT 3341 3378 PLECTIN 48.
FT REPEAT 3379 3416 PLECTIN 49.
FT REPEAT 3420 3454 PLECTIN 50.
FT REPEAT 3794 3836 PLECTIN 51.
FT REPEAT 3837 3874 PLECTIN 52.
FT REPEAT 3875 3912 PLECTIN 53.
FT REPEAT 3913 3950 PLECTIN 54.
FT REPEAT 3954 3988 PLECTIN 55.
FT REPEAT 4328 4370 PLECTIN 56.
FT REPEAT 4371 4408 PLECTIN 57.
FT REPEAT 4409 4446 PLECTIN 58.
FT REPEAT 4447 4484 PLECTIN 59.
FT REPEAT 4488 4522 PLECTIN 60.
FT REPEAT 4862 4904 PLECTIN 61.
FT REPEAT 4905 4942 PLECTIN 62.
FT REPEAT 4943 4980 PLECTIN 63.
FT REPEAT 4981 5018 PLECTIN 64.
FT REPEAT 5022 5056 PLECTIN 65.
FT REPEAT 1935 1951 COILED COIL (POTENTIAL).
FT DOMAIN 2593 2624 COILED COIL (POTENTIAL).
FT DOMAIN 3127 3158 COILED COIL (POTENTIAL).
FT DOMAIN 3651 3692 COILED COIL (POTENTIAL).
FT DOMAIN 4195 4226 COILED COIL (POTENTIAL).
FT DOMAIN 4729 4760 COILED COIL (POTENTIAL).
SQ SEQUENCE 5065 AA; 553078 MW; 378A472DEDA19489 CRC64;

Query Match 3.4%; Score 110.5; DB 1; Length 5065;
Best Local Similarity 21.9%; Pred. NO. 49;
Matches 150; Conservative 80; Mismatches 251; Indels 203; Gaps 36;

QY 28 WHPARTIAGETGTSAPLGLVL--TTPHNISLSPPROLLGFCAEVSGLSTERYELAV 85
DQ 1008 WEQARLLEAVAT-----GGIIDPISHH-----LMPVAIQGY-VDOEMETAL 1052
QY 86 ALAOKNVKLSLEQRLCLARLSEP-----PEDIDALPLDLLFLNDPAFSG 131
DQ 1053 SSSSEFPPTPGOGRTSYAOLEECPREDTSGHLPLPESAPALPTREQVQRSIQAVPG 1112
QY 132 PQACRREFRSRTKAN-VDLLPRG---APEKQRLLPALAC---WGVSGSLSEADVAL 183
DQ 1113 AKDGSIMDLSSCHFTTEQRRGLIEDVQEGRTVYPQLLASVQRWVQETKLLAQARVWP 1172
QY 184 GLIACDLPL-----GRVASEAVLLPRLVSCPPLDDOQOEARALOGGAPPYG--- 233
DQ 1173 GPRG-EYAVAVMLDAGITTOETLELAOGTOS---PAQVADQPAVKALMKGTCAGVLL 1228
QY 234 PPSTWVSVTMDALR-GLLPV-LGQPIIRS-IPQGIIVAMRORSSSDPSWROPERTILPR 290

```



```

Db 1229 QPSGAKASIAQAVNDGLPTGLGQRLLEAQVAGSGLV-----DPLNNG----- 1271
Qy 291 FRRREVENTACSSGKKAREIDESLIFYKKWELEACVDALLATQMDRVNAIFTEQDVL 350
Db 1272 -RLSVE-DAAKVGVLGRVSEQL-----GQERAAA----- 1300
Qy 351 KHKLEDELPQSGPSPSVIOHLGYLEFLKMSPEDIRKWNVTSLETLKALLEVDKGHEMSQAP 410
Db 1301 -----GYPD-----PYSRASLSLMQAM-----ENG-LVPGNE 1326
Qy 411 RRLPQV--AT-LIDRFVG-----RGOLDKDTLDTLA-----FYPGYLC 448
Db 1327 GLPLQVOLAIGVYVD--VHGVLHPQAAACRLGLDPTQSVLTAVDKDKNFEPDPSARD 1385
Qy 449 SLSPSEELSS-----VPPSSIMAVRPOLDICDPQDLQVLYKARLAFQNMMS 496
Db 1386 QVTYQQLRERCVCDSFGLLLPLPSTVLEVDHTVALAMKV--PVSTGRKGSVS 1443
Qy 497 EYFVKIOSFLGAPTEDLKALSQONVMDLATFMKLRITDAVLPITVAEVQRLGPHVEGL 556
Db 1444 LMDLLSEYVADKRRELVALCRSG---RAALRQVYSATVALEAEARQPLQATFRGL 1499
Qy 557 KAERHRPV--RDWILRQRODL-----DTLGLOGGIPNGTL--VLDLSV 599
Db 1500 RKQVSARDLFRALISRKTLDELQGTTVKVEVAMDSVKRSLSBG--NFIAGVLLQGT 1556
Qy 600 QETLSGTPCL-----LGPGVLTVL 619
Db 1557 QERMSIPALRHRILRGTALVLL 1580

```

## RESULT 5

```

DPO1_THECA STANDARD; PRT: 834 AA.
ID DPO1_THECA STANDARD; PRT: 834 AA.
AC P80194;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE I, THERMOSTABLE (EC 2.7.7.7) (TAC POLYMERASE 1).
GN POLA.
OS Thermus aquaticus (subsp. caldophilus).
OC Bacteria; Thermus/Delnooccus group; Thermus group; Thermus.
OX NCBI_TaxID=272;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GK24;
RA Kwon S.-T., Kim J.S., Park J.H., Kim H., Lee D.-S.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RC MEDLINE=93285135; Pubmed=8508785;
RA Park J.H., Kim J.S., Kwon S.-T., Lee D.-S.;
RT "Purification and characterization of Thermus caldophilus GK24 DNA polymerase."
RT Publication and characterization of Thermus caldophilus GK24 DNA polymerase."
RT Eur. J. Biochem. 214:135-140(1993).
CC -1- FUNCTION: THERMOSTABLE ENZYME THAT HAS 5'-TO-3' EXONUCLEASE ACTIVITY AND NO 3'-TO-5' EXONUCLEASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE + DNA(N).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-A FAMILY.
CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements/ or send an email to license@ebi.ac.uk).
CC EMBL, U62584; AAB81398.1;
DR PIR: S33287; S33287.

```

```

DR HS5P; P19821; 17AU.
DR InterPro: IPR002421; 5.3. exonuclease.
DR InterPro: IPR002298; DNA pol.
DR InterPro: IPR001098; DNA pol.
DR InterPro: IPR000513; Exo_N1.
DR InterPro: IPR003583; HNH_1.
DR InterPro: IPR003584; HNH_2.
DR Pfam: PF01367; 5.3. exonuclease; 1.
DR Pfam: PF00476; DNA pol. A; 1.
DR PRINTS: PR00868; DNAPOL1.
DR SMART: SM00475; 53EXOC; 1.
DR SMART: SM00278; Hnh1; 1.
DR SMART: SM00279; Hnh2; 1.
DR SMART: SM00482; POLAC; 1.
DR PROSITE: PS00447; DNA POLYMERASE A; 1.
DR Transferrase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW DNA-binding.
FT DOMAIN 412 834 POLYMERASE (BY SIMILARITY).
SQ SEQUENCE 834 AA; 93798 MW; A851FF3C3076348E CRC64;

```

Query Match 3.48; Score 109.5; DB 1; Length 834;  
 Best Local Similarity 19.48; Pred. No. 4.7; Mismatches 241; Indels 273; Gaps 34;  
 Matches 143; Conservative 80;

```

Qy 48 GLTTPHNISLSPRO-----LIGFPCAEEVSGLSTEREVELAVALAOKNKLSTEQ 98
Db 85 GRAPTEDEF-----PRQALIKELVDLIGFTLVEPGYADV-----LATLAKNEKEGE 136
Qy 99 LRCLAHRLSEPPEDIALPLDLFLNPDAF-----SGPQACTRFSRTIKAN 146
Db 137 VRILT-----ADRDLDQVSDRVAVLHPEGLHITPEMLMOKYGLKEQWDFRALVGDPS 191
Qy 147 VDLPL--RGAPERQRLPALACMCVRCGLSEAD-----VRA-----IG 184
Db 192 -DNLPGVKIGETAL--KLLEWGSLENLKLNDRVKNPVNREKIKHLEDRLSELS 248
Qy 185 GLACDL-----GREY-AESAELVLP-----LVSCGPLODQOQEAARAAL 225
Db 249 RVRDLDPLREVLAQGRERDREGIARFLERLFFGSLHFEGLLEAPARLEAPWP----- 302
Qy 226 QGGGPPV-----PPTWS-----VSTMDLRGL----- 250
Db 303 ---PPEGAFVGFVLSRPEPMMAELKALACRQGVHRAADPLAGLKDLKKEVGRGLADL 358
Qy 251 -----PVLGQPII--RSIQGIYAAMRQSSNRPSRQREERTILRRF 291
Db 359 AVLASREGLDLVPGDDPMLLAVLLDPSWTPPEGVARRRGGEWTEDA--HRLLSERL 414
Qy 292 RR-----EVEK--TACPSGKKAREIDESLIFYKKWELEACVDALLAT 332
Db 415 HRNLKRLQGEKELMLYHEVEKPLSRVLAHMEATGVALDAVYALSLAEELRLEE 474
Qy 333 QMDRVNAIPTYEQLDVLKHL--DEL-----YQGGP--ESV 366
Db 475 EVFRLAGHPFLNSRDQLERVLFDLRLPALGKTOKTGKRSSTAVALBALREAHPIYEKI 534
Qy 367 IOHLGYFLKMSPEDIRKWNVTSLETLKALLEVDKGHEMSQAPRRPLQVATLIDRVK 426
Db 535 LQH-----RELTLKNTYVDPLPSLVH-----PNTGRLLTRF-- 566
Qy 427 GRGLDKDTLDTLTAIFYGYLCSLSEELSSVP-----PSSIMAVRPOLD 472
Db 567 -----NQTATATGRSSSDP--NLQNIPTVTPILGQRIIRAFVAFAGALVALDYS 614
Qy 473 TCDDPQLDVLVPKARLAFQNNNGSEYFVKIOSFLGAPTEDLKALSQONVMDLATEFKL 532
Db 615 QIELRLAHLGSDENLIVFQEGKDIIHTQTSAMGVPPEAVDPLMR--AAKTVNFGVL 672
Qy 533 RTDAVLPITVAEVQKLLGPHVEGLKAERH--RPVADWILRQRODLDTLGLGLOGGI 588
Db 673 YGMSAHRLS---DELAIPEEAFAVIERIFYOSPPKRAVMIKT-----LEGR 717

```





DR SMART; SM00546; CUE; 1.  
 KW SIGNAL: Alternative splicing.  
 FT CHAIN 1 37 POTENTIAL.  
 FT DOMAIN 160 165 ANCIEN UBQUITOUS PROTEIN 1.  
 FT VARSPLIT 114 179 POLY-PHE.  
 FT CONFLICT 354 354 MISSING (IN SHORT ISOFORM).  
 FT SEQUENCE 476 AA; 53028 MW; A6AF3AD84AE86CB CRC64;  
 S -> P (IN REF. 3).  
 Query Match 3.3%; Score 107.5; DB 1; Length 476;  
 Best Local Similarity 22.6%; Pred. No. 3;  
 Matches 114; Conservative 48; Mismatches 157; Indels 185; Gaps 31;  
 QY 52 TP--HNSSLSPPRLGPPCAEVSGLSTERYRELAVAL--AQNNVKISTQLKCLAHRL 106  
 DB 98 TPEFDHNVNMLT-----TCSTVS---ESEAESATGRFPAQAKAPISP-----LAERM 142  
 QY 107 SEPPEDDALPLDLFLFNPDAFSGPACTFFSRTFRKANVDLLPRGAPERQRLT--PAA 164  
 DB 143 ----EDTEALPLTPLY-----PTCOFF--FITLNFILAFSSPSGQPLINSPS 187  
 QY 165 LACGVRGSLLEADYRALGGLADLPGRVAESAENVLLPRLVSCPGFLDODQOEARAA 224  
 DB 188 FVCM--RGFM---EMNGRGELVESLK--RFCASR---LP-----PPLLLPEBEATING 233  
 QY 225 LOGGPPRPSTWSVTMDALRGILPVLGPIIRSTIQGIYAAMROSSRDPSKROPER 284  
 DB 234 REG---LLRFSSWPFSTIQDV-----QPLTLQVGRPLVSV---TVSDASW----- 272  
 QY 285 TILRRFRREVEKTACPSGKKAREIDSLIFYKKWLEACVDALLATQMRVANAIPETY 344  
 DB 273 -----VSELL-----WSL-----FVPEYTV 286  
 QY 345 EQLDVLK--HK---LDLRYOGYPESVYIOLGYLFMKSPEDIRKMWNTSLKALLE 398  
 DB 287 YQVRLRLRVHROLGEANEFEALRVQOLVAKELGCGTRLTAD-----KAENHKKR--- 336  
 QY 399 VDKGEMSPQAPRRPLP-----QVATL-----IDRFYKRGQLD-- 432  
 DB 337 -QRHRLCPQASQSEFPSPGPPSPDVOLATLQRYKEVLLPHVPLGVIGORDLAKTCVULT 395  
 QY 433 -KDTLDTLTAAVPGYLC---SLSPBELSVPPS-----SIWAPRODDLT 473  
 DB 396 ITNLLEGAVAMPEDITGTQSLPTASAKSPSSGPPVTPQPTALTFASMA--RQESLD- 453  
 QY 474 CDPROLVDLYPKARLAFQNNNGSE 497  
 DB 454 --ERKQALYEYARRRRFERRAOE 474  
 RESULT 8  
 OPLA\_RAT STANDARD: PRT; 1288 AA.  
 AC P97608;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE 5-OXOPROLINASE (EC 3.5.2.9) (5-OXO-L-PROLINASE) (PYROGLUTAMASE)  
 DE (5-OPASE).  
 GN OPLAH.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NC NCBL;Taxid-10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Kidney;  
 RX MEDLINE-97113037; PubMed-8943290;  
 RA Ye G.-J., Breslow E., Meister A.;  
 RT "The amino acid sequence of rat kidney 5-oxo-L-prolinase determined  
 by cDNA cloning.";  
 RL J. Biol. Chem. 271:32293-32300(1996).

RN [2]  
 RP ERRATUM.  
 RA Ye G.-J., Breslow E., Meister A.;  
 RL J. Biol. Chem. 272:4646-4646(1997).  
 CC -1- FUNCTION: CATALYZES THE CLEAVAGE OF 5-OXO-L-PROLINE TO FORM L-  
 CC GLUTAMATE COUPLED TO THE HYDROLYSIS OF ATP TO ADP AND INORGANIC  
 CC PHOSPHATE.  
 CC -1- CATALYTIC ACTIVITY: ATP + 5-OXO-L-PROLINE + 2 H(2)O = ADP +  
 CC PHOSPHATE + L-GLUTAMATE.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- TISSUE SPECIFICITY: WELL EXPRESSED IN TESTIS, KIDNEY AND LIVER.  
 CC -1- SIMILARITY: BELONGS TO THE OXOPROLINASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL; U70825; AAC52955.1; -  
 DR InterPro: IPR002821; Hydatoinase\_A.  
 DR InterPro: IPR003692; Hydatoinase\_B.  
 DR Pfam: PF01968; Hydatoinase\_A; 1.  
 DR Pfam: PF02538; Hydatoinase\_B; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 1288 AA; 137746 MW; 4530D0781E10C7AD CRC64;  
 Query Match 3.3%; Score 107.5; DB 1; Length 1288;  
 Best Local Similarity 20.4%; Pred. No. 12;  
 Matches 140; Conservative 69; Mismatches 224; Indels 253; Gaps 34;  
 QY 19 LFLFLSLGWVHAKRLTAGETGSAPLGVL---TPPHNSSLSPROLGFPAAEVSGL 75  
 DB 265 VLFMR-SDGGLAPMDAFSAVSGRAVSGVAGVYATYHLEGCP--VIGF--DMGT 318  
 QY 76 ST-----ERYRELAVALAQNNVKISTQLKCLAHRLSEPPEDDALPL 118  
 DB 319 STDVSRVAGEFEHVEFESTA---GVTLQAPOL-----DITVAAGGSRLEFF 362  
 QY 119 -DLLFLNPD---AFSGPACTREFSRITKANVDLLPRGAPERQRLPALACGVRGSL 174  
 DB 363 RSGLFVVGPEESAGAHGP-ACYRKGGPVTVADNLV-----LGRLLPASPCC----- 408  
 QY 175 LSEADYRALGGLADLPGRVAESAENVLLPRLVSCPGFLDODQOEARAAALQGGPP- 232  
 DB 409 -----IFGPGEDQPLSPESASRKALEAVAMEVNS 436  
 QY 233 ---GPPSTWSVTMDALRGILPVLGPIIRSTIQGIYAAMROSSRDPSKROPERITLR 288  
 DB 437 FLTNGPCPASQSLSEAVAMGEFVYANENACRP-----RALTOARGHDS----- 481  
 QY 289 PRFRREVEKTACPSG-----KKAREIDESLIFYKKWLE---ACVDALLATQMD 335  
 DB 482 -----AHVLACFGAGGQHACATARALGMDTVHHRHSGLSALGLADLVNHEADQPC 535  
 QY 336 RVNAIPETFEQDLVAKHLDE---LYPGYPESVYIOLGYLFMKSPEDIRKMWNTSL 390  
 DB 536 SLSTYPERFAQLDQRLSLREQCVDALVOGFPSPQISTESTESFLHRYOGTCC----- 587  
 QY 391 ETIKALLEVDGHEMSPPQAPRRP-----LPOVATLIDRE-VKGRG--- 429  
 DB 588 ---ALWYSAHQHPATACSPRAGDGAAFVRYKAREFGFIIPERVVVDVAVKGTGNSG 643  
 QY 430 -QLDKDT-----LDTLTAFY-----PGYL----- 447  
 DB 644 LQLE-DTPKIGTGPVHVKVQCYFEGGYQETPVYLLGELGYGHQLOGPCLIDNNSTIL 702  
 QY 448 ---CSLSPBELSVPPSIAVNRPODDLT--CDPROLDVLYPKARLAFQNNNGSEYFKI 502  
 DB 703 VEPGQAEVTDGTGDI-RISVGAEGPSMADTRLDPIOLISF-----SHREMSI 748

QY 503 OSFLGAPTEDEKALSOONVSMDLATFMKLRDAVLPLTVAEVOKILGPHVEGLKAEEH 562  
 DB 749 AEQMG-----RILORNAIS-----TNKEREDFSC-----ALFGPD-GGLVSNAPH 788  
 OY 563 RPYRWILROR-QDDLDPLGLGLOGG 587  
 DB 789 IPVHLGAMQETVOFOIHLGADLHPG 814

RESULT 9  
 ERY2\_SACER STANDARD. PRT: 3567 AA.  
 AC 003132; Q54096;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE ERYTHRONOLIDE SYNTHASE, MODULES 3 AND 4 (EC 2.3.1.94) (ORF 2) (6-  
 DEOXYERYTHRONOLIDE B SYNTHASE II) (DEBS 2).  
 GN ERYA.  
 OS Saccharopolyspora erythraea (Streptomyces erythraeus).  
 CC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:  
 CC Actinomycetales: Pseudonocardiaceae; Pseudonocardiaceae;  
 CC Saccharopolyspora.  
 CC NCBI\_TaxID=1836;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91220065; PubMed=2024119;  
 RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;  
 RT "Modular organization of genes required for complex polyketide  
 RT biosynthesis.";  
 RL Science 252:675-679(1991).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL 2338;  
 RX MEDLINE=92155230; PubMed=1740151;  
 RA Bevilitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;  
 RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.  
 RT Cloning of the structural gene, sequence analysis and inferred domain  
 RT structure of the multifunctional enzyme.";  
 RL Eur. J. Biochem. 204:39-49(1992).  
 CC -1 CATALYTIC ACTIVITY: 6 METHYLMALONYL-COA + PROTONYL-COA -> 7 COA  
 CC + 6-DEOXYERYTHRONOLIDE B.  
 CC -1 COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.  
 CC -1 PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN  
 CC BIOSYNTHESIS.  
 CC -1 MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH  
 CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3  
 CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH  
 CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS  
 CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,  
 CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,  
 CC RESPECTIVELY.  
 CC -1 MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),  
 CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACTYL CARRIER  
 CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),  
 CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE  
 CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION  
 CC OF THE FULL-LENGTH CHAIN.  
 CC -1 SIMILARITY: TO FATTY ACID SYNTHASE (FAS).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: M63677; AAA26494.1;  
 DR EMBL: X62569; CAA44448.1;  
 DR HSSP: P28304; 100R.  
 DR InterPro: IPR001227; Acyltransf\_domain.

DR InterPro: IPR002085; Adh\_zinc.  
 DR InterPro: IPR000794; Ketoacyl-synt.  
 DR InterPro: IPR003880; Phosphopantl\_attach.  
 DR Pfam: PF00698; Acyl\_transf; 2.  
 DR Pfam: PF00107; adh\_zinc; 1.  
 DR Pfam: PF00109; ketoacyl-synt; 2.  
 DR Pfam: PF00550; pp-binding; 2.  
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.  
 DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE; 2.  
 DR PROSITE: PS0075; ACP\_DOMAIN; 2.  
 DR Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;  
 DR Phosphopantetheine; Multifunctional enzyme.  
 KW DOMAIN 1 1484  
 FT DOMAIN 1 1485 3567  
 FT DOMAIN 2 27 488  
 FT DOMAIN 559 884  
 FT DOMAIN 1130 1301  
 FT DOMAIN 1397 1467  
 FT DOMAIN 1485 1943  
 FT DOMAIN 2013 2336  
 FT DOMAIN 2383 3066  
 FT DOMAIN 3139 3322  
 FT DOMAIN 3415 3485  
 FT ACT\_SITE 202 202  
 FT ACT\_SITE 651 651  
 FT BINDING 1430 1430  
 FT ACT\_SITE 1661 1661  
 FT ACT\_SITE 2115 2115  
 FT NP\_BIND 2961 2978  
 FT NP\_BIND 3142 3157  
 FT BINDING 3448 3448  
 FT CONFLICT 438 438  
 FT CONFLICT 480 480  
 FT CONFLICT 1241 1241  
 FT CONFLICT 2664 2664  
 SO SEQUENCE 3567 AA; 374413 MW; EE6284F4738AA0C0 CRC64;

Query Match 3.3%; Score 107.5; DB 1; Length 3567;  
 Best Local Similarity 20.4%; Pred. No. 49;  
 Matches 147; Conservative 63; Mismatches 225; Indels 285; Gaps 35;

QY 10 WSCGDPESLFLFLSLGCVHPARTLAGEITGESAPLGLGVITTPNINISLSPRLGLGPC 69  
 DB 1885 WPAGERP-----RAGVSSGCVG--TNAIVYENP----- 1914  
 QY 70 AEVSGLSERVELEAVALAOKNVKLSTEOQLCLAHRLSEPPEDLDALPLDLLFLNPDAF 129  
 DB 1915 AEQEARTER-GPLPFLVLSGRSEAVVAQAARALAEHLRDTPE-----LGLTDAAM 1963  
 QY 130 SGPAQCTEFPSRITK-----ANVDLLPRGAPERRLPALACGVCGLSLEADY 180  
 DB 1964 TLATGRARFDVRAAVLGDRAVCACELDALAEGRSADAVAPVISA-----PRRPV 2014  
 QY 181 RALGLACDLPG--RFVSAEVLPLRLVSCPGPLDQDOEAAARAALOGGPGPYGPSTW 238  
 DB 2015 LVFPGQAGWGMARNLDESSEVFAESMSKCAEL-----SPHIDW 2055  
 QY 239 SVSTMDALRG-----LLPVLGOPITRSTPGIVAAAMRORSSRDSRPOPERTILPR 290  
 DB 2056 KL--LDVVRGGGDPDPRHRVVL-QPVLSIMVSLAELMRAHG-----VTP- 2098  
 QY 291 FRREVEKACPSGKKARIDESTLFYKKWELEACVDALLATMDRVNAIPFTYEQDVL 350  
 DB 2099 -----AAVYGHSGGITAIAHV-AGALSLDAAKVALNSQV----- 2133  
 QY 351 KHKIDELYPQGPSPSVIOHLGYLEFLKMSPEDIRKWNVTSLETLKA----- 395  
 DB 2134 ---LRELDQG-----GMVSGASRDE-----LETVLARKMGRAVAANVPG 2173  
 QY 396 -----LLEVD-----KGHEKSPQ--APR--RPLPOVALTIDRFYKGGOLDKDTLDT 438

```

Db 2174 TSVACPTAEIDEEFAAEAREMKPRRIAVRYASHSEPARIEDRLAA-----ELGT 2225
Qy 439 LTAIVPGVGLSPEELSSVPSIMAVRPDLTCDPRDVLVYPRARLAFQNMNGSEY 498
Db 2226 ITAV-----RGSVPLHS--TVTGEVDT--SAMDASY-----WYRNLRRPVL 2263
Qy 499 FVKIQSEFGAPTEDLKALSOQNVSMDLATPMKRLRTDAVLPLTVAE-----544
Db 2264 F-----EQAIVGVQGF-----TFVEVSPHVLVMAVEETREHGAETCVPT 2308
Qy 545 -----VQKLGPHEVGLAEER-----HRPV-----RDWILRQRODDL 578
Db 2309 LRREOSGHEERLRLRAHVGVGADLRPAVAGRPALPTPYPEHOFMRPRPADVS 2368
Qy 579 TLGT-----GLGGIPNGVLYLDS--VQETLSGPPCLLGGPVLTVL 621
Db 2369 ALVGRAEHPHLLAAVDVPGHGAFTGRSLSTDQAPLAEHVGGRTLV--PGSVLVDL 2427

RESULT 10
APPL_MOUSE STANDARD; PRT: 653 AA.
AC 003157;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE AMYLOID-LIKE PROTEIN 1 PRECURSOR (APLP).
GN APLP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9306322; PubMed=1279693;
RA Masco W., Bupp K., Magendantz M., Gussella J.F., Tanzi R.E.,
RA Solomon F.;
RT Identification of a mouse brain cDNA that encodes a protein related
RT to the Alzheimer disease-associated amyloid beta protein precursor.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:10758-10762(1992).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. MAY BE PROCESSED
CC IN THE GOLGI COMPLEX; HOWEVER, IT IS NOT YET CLEAR WHETHER APLP
CC IS SECRETED.
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL: L04538; AAA37247.1; -
CC PIR: A46362; A46362.
CC MGI: MGI:88046; APLP1.
CC InterPro: IPR001868; A4_APP.
CC Pfam: PF02177; A4_EXTRA.1.
CC SMART: SM00006; A4_EXTRA.1.
CC PROSITE: PS00319; A4_EXTRA.1.
CC PROSITE: PS00320; A4_INTRA.1.
CC Glycoprotein; Transmembrane; Signal.
KW SIGNAL
FT CHAIN 1 20 AMYLOID-LIKE PROTEIN 1.
FT DOMAIN 21 653 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 584 606 POTENTIAL.
FT DOMAIN 607 653 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 643 646 CLATHRIN-BINDING (POTENTIAL).
FT DOMAIN 263 271 POLY-GLU.
FT CARBOHYD 464 464 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 554 554 N-LINKED (GLCNAc. . .) (POTENTIAL).

```

```

SQ SEQUENCE 653 AA; 72751 MW; 56516DC3E4A0E4B0 CRC64;
Query Match 3.38; Score 106.5; DB 1; Length 653;
Best Local Similarity 20.28; Pred. No. 5.4;
Matches 119; Conservative 72; Mismatches 220; Indels 179; Gaps 27;
Qy 135 CTREFSRTKANYDLPRGAPERORLPAALACWGVSGSLSEADVRLGLACDLPRF 194
Db 94 CRMPPELHIAKVE-----QAAQALPMERNCGSTSGRAHHHEVY--PFGC-LPGEF 144
Qy 195 VAESAELLP-----RLVSCGPRL--DQDQEAARA--ALGGP--PYG-----233
Db 145 VSEA--LLVPEGCRFLHQRMDQCESTRHQEAQCSOGILLHSGMILLPGSDRFR 202
Qy 234 -----PESTSVSTMDALRGILLPVLCOPTIRSIQGIYAAMAROSSRRPSRPF-ER 284
Db 203 GVEYVCCPPATPNPSSGAA-----GDPSTRSWPLGGRAEGDEEVESEFPQVDD 254
Qy 285 TILRRPREVEKTPACPSGKKAREIDSLFYKKMLEACVDALLATQMDRNAIPFTY 344
Db 255 YFEP-----FOAEDEEEDEE-----RAPPSHTPYMSRYVTPTRPT 294
Qy 345 EQLDVLKHKIDELYPQYPSVIGHLYGLKMSPEDIRKWNVTSLTKALLEVDKGE 404
Db 295 DGVDVY-----FGMPGEIGHEGFLRAKMDLEERRMOIN--EYRMAMADSQSK 343
Qy 405 MSPQAPRRPLPQ-----VATLIDRFYKRGQIDKDT-----LDLTATPYGLDSL 451
Db 344 NLPRARQALNEHFOSLIQTLEQVSGERQLVETHATRYALINDORRALEGFALAQ 403
Qy 452 P-----ELSSVPSSSIMAVR-----467
Db 404 GDPQAEKRVMLALRYLAEQKQRIHLRIHQHVAANDPEKAQOMRQOVQTHLVIERM 463
Qy 468 PDLDTCD--PRDLVLYPKAR--LAFQNMNGSEYFVKIQSEFGAPTEDLKALSOQNV 523
Db 464 NQSLGLDQNHLAGLELRFQICELLAEHLGPSE-----LDASVPGSSSEDKSLQAPESK 519
Qy 524 MD-LATFMKLTLD-----AVLPLVAIVOKILGPHVSGLAEEHNRVRYMILQR 573
Db 520 DDPVTLPGKSTDOBSSSGEKKLPRLQYE-QKVNASAPRPFPHSSD-----IQ 569
Qy 574 QDDLDLTGLGGIPNGVLYLDSVQETLSGTPCLLGGPVLTVLALL 623
Db 570 RDELAPSGTGS-----REALSGLLIMGGGSLYLSSL 605

RESULT 11
HRPO_BURSO STANDARD; PRT: 690 AA.
ID HRPO_BURSO
AC P35656;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPERSENSITIVITY RESPONSE SECRETION PROTEIN HRPO.
GN HRPO.
OS Burkholderia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_Taxid=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RC MEDLINE=93302711; PubMed=8316211;
RA Gough C.L., Genin S., Lopes V., Boucher C.A.;
RT "Homology between the hrpo protein of Pseudomonas solanacearum and
RT bacterial proteins implicated in a signal peptide-independent
RT secretion mechanism.";
RL Mol. Gen. Genet. 239:378-392(1993).
CC -1- FUNCTION: INVOLVED IN THE SECRETION OF A PROTEINOGENOUS ELICITOR OF
CC THE HYPERSENSITIVITY RESPONSE IN PLANTS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE

```





```

CC -1- CATALYTIC ACTIVITY: ATP + L-GLYCINE + TRNA(GLY) = AMP +
CC PYROPHOSPHATE + L-GLCYL-TRNA(GLY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U20547; AAC62982.1; -
DR EMBL: AE001352; AAC68391.1; -
DR Interpro: IPR002106; AA_TRNA_Ligase_II.
DR Interpro: IPR002310; tRNA_synth_2e.
DR Interpro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02091; tRNA_synth_2e; 1.
DR Pfam: PF02092; tRNA_synth_2f; 1.
DR PRINTS: PRO1044; TRNASYNTHGA.
DR PRINTS: PRO1045; TRNASYNTHGB.
DR PROSITE: PS00179; AA_TRNA_LIGASE_II_1; FALSE,NEG.
DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; 1.
DR K1 Aminoacyl-tRNA synthetase: Protein biosynthesis; Ligase; ATP-binding;
KM Complete proteome.
FT DOMAIN 1 310 GLCYL-TRNA SYNTHETASE ALPHA CHAIN.
FT 103 1003 GLCYL-TRNA SYNTHETASE BETA CHAIN.
FT VARIANT 10 103 T -> A (IN SEROVAR L2).
FT VARIANT 186 186 D -> G (IN SEROVAR L2).
FT VARIANT 221 221 F -> S (IN SEROVAR L2).
FT VARIANT 405 405 S -> P (IN SEROVAR L2).
FT VARIANT 424 424 W -> C (IN SEROVAR L2).
FT VARIANT 565 565 I -> V (IN SEROVAR L2).
FT VARIANT 704 704 T -> P (IN SEROVAR L2).
FT VARIANT 738 738 S -> N (IN SEROVAR L2).
FT VARIANT 838 838 A -> E (IN SEROVAR L2).
FT VARIANT 864 864 G -> I (IN SEROVAR L2).
FT VARIANT 1001 1001 V -> D (IN SEROVAR L2).
FT CONFLICT 602 618 ORFEPNOMGCEITNRF -> PIVLPYTKYARRNHQSI
FT (IN REF. 1).
FT S -> SARRAS (IN REF. 1).
FT CONFLICT 828 828 MY -> TH (IN REF. 1).
FT CONFLICT 902 903
FT SEQUENCE 1003 AA; 112576 MW; 79EB918FE19D39E9 CRC64;

```

Query Match 3.28; Score 104; DB 1; Length 1003;  
 Best Local Similarity 21.48; Pred. No. 15;  
 Matches 130; Conservative 72; Mismatches 221; Indels 184; Gaps 34;

```

QY 38 ETGESAPLGVLTTPHNISLSPROL---LGFPCAEGVSGLSTERYRELAVALAOKNVK 93
QY 318 EIGSEELPAFVFPTGIGQOLESIAKKLADHGIAYKHLEVGTP---RRALAC----- 366
QY 94 LSTQQLCLARLSEPPEDDALPLDLLFLNPAFSGPQACTFFSRITRANVDLPK 153
QY 367 -----IGLSIVTIRPSEKKGPLSLFLM---TIGSVSPQG-EOGFFS---SHGISISHRS 415
QY 154 APERQRLPALACMGVRSGLSADVRAAGLACDLPGRVVASAEVL---LPRVSC- 209
QY 416 ALDO---PSKI---WRKR-----SINGIDYLFVLVPEE-RKETAILVVELQQLINSI 461
QY 210 --PGPLDODOEAAARALOGGPGPYGPSTWSVTMDALRG--LLPYLGOPPIRSIPOGI 265
QY 462 RFPQKMMWD-----NGVEYARPIRWLV---ALYDQQLP-----ISLGF 498
QY 266 VAAAROSSRDPSV--RO-PERTILRPFREREVE--KTACPSGKKAREIDSLIFYKWE 320
QY 499 V-----SSGWTSGHQLDRQLTIPSSNMKYVDTLRSAC-----YIVSQKE 539
QY 321 LEACVDAALLATQMDRNVATP-----FTYQOLDVYLKKIKLDELVPQGYPE---SVI 367
QY 540 RRAIIKQGLQNLTDQIVAIAPRHLIDETVFLTEHPVIAISOFPARCSLPKEILLIAMI 599

```

```

QY 368 OHIGYFLKMSPEDIRKWNYSLETLEKALLEVD-----KGHEMSPOAPRRRLPOVAT 419
QY 600 QHQRIF-----PTNOMGCEITN-----RFLIVCDNSPDISSVEGNEKA----- 637
QY 420 LIDRFVKGQGLDKDLDLTLPATFAPGYGLCSLPEELSSVPPSS-----IWAVRP-- 468
QY 638 LAPRLTGNGNLFKODLLTLPSSFVEKLSVYTFESLGLSLADKTRSLKHLAEAVALLPLC 697
QY 469 --QDLTDCRDOLDVLYPKARLAFQNMNGSEYFYKISFPGAPTEDELKALSOQNVSMDL 526
QY 698 AKEDIDTA-----IHYCKADLVSSVVA---EPEELQING-----RYIQNASLSR 740
QY 527 ATFW-----KLRTDAVLPLTVAEVOKLGLPHVEGLKAEERHRRVPMDILROR 573
QY 741 AALAIIEHLOHTITGSSISTTGALLSIDRIDNLSLCSFILGLTPSSHP---YALRRQ 797
QY 574 ODDLDLT 580
QY 798 SLEITL 804

```

RESULT 15  
 SPK\_HUMAN STANDARD; PRT; 1142 AA.  
 AC Q92797; O00733; O00689;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE STUPLEKIN.  
 GN SPK.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND MASS SPECTROMETRY.  
 RC TISSUE=Colon carcinoma;  
 RX MEDLINE=96354916; Pubmed=8769423;  
 RA Kaon B.H., Schaefer S., Kuhn C., Grund C., Franke W.W.;  
 RT "Symplekin, a novel type of tight junction plaque protein.";  
 RL J. Cell Biol. 134:1003-1018(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RX MEDLINE=98252805; Pubmed=9585442;  
 RA Alwazzan M., Hamshire M.G., Lennon G.G., Brook J.D.;  
 RT "Six transcripts map within 200 kilobases of the myotonic dystrophy  
 expanded repeat.";  
 RL Mamm. Genome 9:485-487(1998).  
 CC -1- FUNCTION: SPECIFIC COMPONENT OF THE TIGHT JUNCTION (TJ) PLAQUE,  
 CC BUT MIGHT NOT BE AN EXCLUSIVELY JUNCTIONAL COMPONENT. MAY HAVE A  
 CC HOUSE-KEEPING RULE.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF ADHESION PLAQUES (MAJOR)  
 CC AND NUCLEOLAR PLASM (MINOR) (IN CELLS WITH TJ). NUCLEOLAR PLASM (IN CELLS  
 CC WITHOUT TJ).  
 CC -1- TISSUE SPECIFICITY: IN TESTIS, EXPRESSED IN POLAR EPITHELIA AND  
 CC SERTOLI CELLS BUT NOT IN VASCULAR ENDOTHELIA. THE PROTEIN IS  
 CC DETECTED IN STOMACH, DUODENUM, PANCREAS, LIVER, FETAL BRAIN,  
 CC CARCINOMAS, LENS-FORMING CELLS, FIBROBLASTS, LYMPHOCYTES, LYMPHOMA  
 CC CELLS, ERYTHROLEUKEMIA CELLS BUT NOT IN ENDOTHELIUM OF VESSELS,  
 CC EPIDERMIS, INTERCALATED DISKS, PURKINJE FIBER CELLS OF THE HEART  
 CC AND LYMPH NODE.  
 CC -1- DIFFERENTIAL DIAGNOSIS OF TUMORS.  
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 09:48:17 ; Search time 55.89 Seconds  
(without alignments)  
1643.568 Million cell updates/sec

Title: US-09-215-035-2  
Perfect score: 3261  
Sequence: 1 MALGRDPCWSCGDRPGSL.....LLGPPVTLVALLASTIA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_17:\*  
2: SP\_Archaea:\*  
3: SP\_Bacteria:\*  
4: SP\_Fungi:\*  
5: SP\_Human:\*  
6: SP\_Invertebrate:\*  
7: SP\_Mammal:\*  
8: SP\_mhc:\*  
9: SP\_Organelle:\*  
10: SP\_Phage:\*  
11: SP\_Plant:\*  
12: SP\_Podent:\*  
13: SP\_Virus:\*  
14: SP\_Vertebrate:\*  
15: SP\_Unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3110	95.4	630	4 Q9BR17	Q9BR17 homo sapien
2	3055	93.7	622	4 Q14859	Q14859 homo sapien
3	3042.5	93.3	621	4 Q9BTR2	Q9BTR2 homo sapien
4	1710	52.4	625	11 Q9ERA7	Q9ERA7 mus musculu
5	1707	52.3	625	11 Q61468	Q61468 mus musculu
6	1526	46.8	698	4 Q9UK57	Q9UK57 homo sapien
7	133	4.1	698	4 Q9FIJ5	Q9FIJ5 pseudomonas
8	131.5	4.0	797	4 Q9UGS3	Q9UGS3 homo sapien
9	126	3.9	939	4 Q9WJX0	Q9WJX0 corynebacte
10	125.5	3.8	2109	8 Q32836	Q32836 pelargonium
11	124	3.8	1906	4 Q15029	Q15029 homo sapien
12	124	3.8	2011	4 Q9P0A9	Q9P0A9 homo sapien
13	120	3.7	2218	10 Q9SL26	Q9SL26 arabidopsis
14	117.5	3.6	3574	10 Q9AUB4	Q9AUB4 arabidopsis
15	116.5	3.6	2015	5 Q9USY1	Q9USY1 dictyosteli
16	116	3.6	1313	4 Q75033	Q75033 homo sapien
17	116	3.6	1429	5 Q9W5A4	Q9W5A4 drosophila
18	115	3.5	4247	2 Q9L8H4	Q9L8H4 streptomyce
19	114.5	3.5	756	2 Q9A888	Q9A888 caulobacter

20	114	3.5	1067	5 Q9Y1P2	Q9Y1P2 drosophila
21	113.5	3.5	817	5 Q9NK95	Q9NK95 drosophila
22	113.5	3.5	1012	4 Q95712	Q95712 homo sapien
23	112.5	3.4	730	6 Q9TUC2	Q9TUC2 equus caball
24	112.5	3.4	1012	4 Q9UKV7	Q9UKV7 homo sapien
25	111.5	3.4	613	12 Q9W8X7	Q9W8X7 porcine ade
26	111.5	3.4	4957	4 Q14687	Q14687 homo sapien
27	111.5	3.4	5262	4 Q14686	Q14686 homo sapien
28	111	3.4	1596	13 Q918E1	Q918E1 fuigu rubrip
29	111	3.4	3851	4 Q431E1	Q431E1 homo sapien
30	111	3.4	3926	4 Q9UPA5	Q9UPA5 homo sapien
31	110.5	3.4	340	10 Q22706	Q22706 arabidopsis
32	110	3.4	912	2 Q9K3Y2	Q9K3Y2 streptomyce
33	110	3.4	1770	2 Q50857	Q50857 myxococcus
34	109	3.3	1171	2 Q9XAH6	Q9XAH6 streptomyce
35	108.5	3.3	636	11 Q9DOC6	Q9DOC6 mus musculu
36	108.5	3.3	652	2 Q9HTA6	Q9HTA6 pseudomonas
37	108.5	3.3	1398	3 Q9URU2	Q9URU2 schizosacch
38	108.5	3.3	2361	5 Q94226	Q94226 caenorhabdi
39	108	3.3	283	2 Q9PD02	Q9PD02 xyella fas
40	108	3.3	793	4 Q9H5Z2	Q9H5Z2 homo sapien
41	107.5	3.3	757	2 Q9AJN7	Q9AJN7 athrobacte
42	107	3.3	451	2 Q9A844	Q9A844 caulobacter
43	107	3.3	1715	6 Q9GLM4	Q9GLM4 bos laurus
44	106	3.3	391	2 Q9Z360	Q9Z360 streptomyce
45	106	3.3	527	2 Q9RWV6	Q9RWV6 delnococcus

## ALIGNMENTS

RESULT 1  
Q9BR17 PRELIMINARY: PRT: 630 AA.  
ID Q9BR17  
AC Q9BR17  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DR 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DN C335H7.1 (MESOTHELIN).  
GN MSIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Codley V.  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL031258; CAC37289.1; -.  
SQ SEQUENCE 630 AA; 68985 MW; FA17E3609B6CC9CA CRC64;

Query Match	Best Local Similarity	95.4%;	Score 3110;	DB 4;	Length 630;
Matches 608;	Conservative	4;	Mismatches 16;	Indels 2;	Gaps 2;
QY 1	MALGRDPCWSCGDRPGSL				
DB 1	MALGRDPCWSCGDRPGSL				
QY 59	LSPROLGFPCAEEVSGSL				
DB 61	LSPROLGFPCAEEVSGSL				
QY 119	DLLFLNPDATFSGPACR				
DB 121	DLLFLNPDATFSGPACR				
QY 179	DYRALGGLACDPGRFVSA				
DB 181	DYRALGGLACDPGRFVSA				
QY 239	SVSTMALRGLLPVGLGPI				
DB 240	SVSTMALRGLLPVGLGPI				

```

|||||
Db 241 SVSTMDALRGLLPVLGOPILRSIPOGIVAAMRORSSRDPSPWOPERTILRPFRREREK 300
QY 299 ACPGSKKAREIDESLIFKKWLEACVDAALLATQMDRVNAPFYEEDLVKHKHLDLXL 358
Db 301 ACPGSKKAREIDESLIFKKWLEACVDAALLATQMDRVNAPFYEEDLVKHKHLDLXL 360
QY 359 POGYVESYIQHLYGFLKMSPEIDIKMNVTSLETALKALLEVDKHEMSPOAPRRRLPOVA 418
Db 361 POGYVESYIQHLYGFLKMSPEIDIKMNVTSLETALKALLEVDKHEMSPOAPRRRLPOVA 420
QY 419 TLIDRFVKGROLDKDTLDTLTAFFPGYLCISLSPBELSSVPPSSIMAVRPODLTCDPRQ 478
Db 421 TLIDRFVKGROLDKDTLDTLTAFFPGYLCISLSPBELSSVPPSSIMAVRPODLTCDPRQ 480
QY 479 LDVLYPKARLAFQNMNGSEYFVKIOSFLGAPTEEDLKALSOQNVSMDLATEPKKLTDAVL 538
Db 481 LDVLYPKARLAFQNMNGSEYFVKIOSFLGAPTEEDLKALSOQNVSMDLATEPKKLTDAVL 540
QY 539 PLTVAEVOKLLGPHVEGLKAEERHNRVDMILRQRDDLDLTGLGLOGGIPNGYLVLDLS 598
Db 541 PLTVAEVOKLLGPHVEGLKAEERHNRVDMILRQRDDLDLTGLGLOGGIPNGYLVLDLS 600
QY 599 VOETLSGTPCLLGGPVLVTALLASTLA 628
Db 601 VOETLSGTPCLLGGPVLVTALLASTLA 630

```

```

RESULT 2
ID 014859 PRELIMINARY; PRT; 622 AA.
AC 014859;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PRE-PRO-MEGAKARYOCYTE POTENTIATING FACTOR PRECURSOR.
GN MPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95394969; PubMed=7665620;
RA Kojima T., Oh-peda M., Hattori K., Taniguchi Y., Tamura M., Ochi N.,
RA Yamaguchi N.;
RT "Molecular cloning and expression of megakaryocyte potentiating factor
cDNA."
RL J. Biol. Chem. 270:21984-21990(1995).
DR EMBL; D49441; BAA08419.1;
DR InterPro; IPR000585; Hemoexin.
DR PROSITE; PS00024; HEMOEXIN; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 286 MEGAKARYOCYTE POTENTIATING FACTOR.
SQ SEQUENCE 622 AA; 68037 MW; 1F0FE52883CA6C13 CRC64;

```

```

Query Match 93.7%; Score 3055; DB 4; Length 622;
Best Local Similarity 95.4%; Pred. No. 4.8e-226;
Matches 601; Conservative 3; Mismatches 16; Indels 10; Gaps 3;
QY 1 MALORLDP-CWSCGDRP-GSLILFLFSLGWHPARTLAGETGESAPLGVLTPPHNISS 58
Db 1 MALPAPRLGSCGCPALGSLILFLFSLGWVQSRITLAGETGEAAPLDGYLANPNISS 60
QY 59 LSPRLGFCPEAVESGLSTERYRELAVALAOKNVKLSLEQLRCLAHRLSEPPEDDALPL 118
Db 61 LSPRLGFCPEAVESGLSTERYRELAVALAOKNVKLSLEQLRCLAHRLSEPPEDDALPL 120
QY 119 DLLFLFNDASSGPOACTRFRFSRITKANVDLLPRGAPRORLLPAAALCKWYRGSLSEA 178
Db 121 DLLFLFNDASSGPOACTRFRFSRITKANVDLLPRGAPRORLLPAAALCKWYRGSLSEA 180

```

```

QY 179 DVRALGLACDLPGRFVAESAEEVLLPRLVSCPGPLDODQOEAARALOGGPPYPSTW 238
Db 181 DVRALGLACDLPGRFVAESAEEVLLPRLVSCPGPLDODQOEAARALOGGPPYPSTW 240
QY 239 SVSTMDALRGLLPVLGOPILRSIPOGIVAAMRORSSRDPSPWOPERTILRPFRREREK 298
Db 241 SVSTMDALRGLLPVLGOPILRSIPOGIVAAMRORSSRDPSPWOPERTILRPFRREREK 300
QY 299 ACPGSKKAREIDESLIFKKWLEACVDAALLATQMDRVNAPFYEEDLVKHKHLDLXL 358
Db 301 ACPGSKKAREIDESLIFKKWLEACVDAALLATQMDRVNAPFYEEDLVKHKHLDLXL 360
QY 359 POGYVESYIQHLYGFLKMSPEIDIKMNVTSLETALKALLEVDKHEMSPOAPRRRLPOVA 418
Db 361 POGYVESYIQHLYGFLKMSPEIDIKMNVTSLETALKALLEVDKHEMSPOAPRRRLPOVA 420
QY 419 TLIDRFVKGROLDKDTLDTLTAFFPGYLCISLSPBELSSVPPSSIMAVRPODLTCDPRQ 478
Db 421 TLIDRFVKGROLDKDTLDTLTAFFPGYLCISLSPBELSSVPPSSIMAVRPODLTCDPRQ 480
QY 479 LDVLYPKARLAFQNMNGSEYFVKIOSFLGAPTEEDLKALSOQNVSMDLATEPKKLTDAVL 538
Db 481 LDVLYPKARLAFQNMNGSEYFVKIOSFLGAPTEEDLKALSOQNVSMDLATEPKKLTDAVL 540
QY 539 PLTVAEVOKLLGPHVEGLKAEERHNRVDMILRQRDDLDLTGLGLOGGIPNGYLVLDLS 598
Db 541 PLTVAEVOKLLGPHVEGLKAEERHNRVDMILRQRDDLDLTGLGLOGGIPNGYLVLDLS 600
QY 599 VOETLSGTPCLLGGPVLVTALLASTLA 628
Db 593 VOETLSGTPCLLGGPVLVTALLASTLA 622

```

```

RESULT 3
ID 09BTR2 PRELIMINARY; PRT; 621 AA.
AC 09BTR2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE MESOTHELIN.
GN MESOTHELIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=PANCREATIC ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003512; AAH03512.1;
SQ SEQUENCE 621 AA; 67908 MW; 1A7367740956997B CRC64;

```

```

Query Match 93.3%; Score 3042.5; DB 4; Length 621;
Best Local Similarity 95.2%; Pred. No. 4.3e-225;
Matches 600; Conservative 4; Mismatches 15; Indels 11; Gaps 4;
QY 1 MALORLDP-CWSCGDRP-GSLILFLFSLGWHPARTLAGETGESAPLGVLTPPHNISS 58
Db 1 MALPAPRLGSCGCPALGSLILFLFSLGWVQSRITLAGETGEAAPLDGYLANPNISS 59
QY 59 LSPRLGFCPEAVESGLSTERYRELAVALAOKNVKLSLEQLRCLAHRLSEPPEDDALPL 118
Db 60 LSPRLGFCPEAVESGLSTERYRELAVALAOKNVKLSLEQLRCLAHRLSEPPEDDALPL 119
QY 119 DLLFLFNDASSGPOACTRFRFSRITKANVDLLPRGAPRORLLPAAALCKWYRGSLSEA 178
Db 120 DLLFLFNDASSGPOACTRFRFSRITKANVDLLPRGAPRORLLPAAALCKWYRGSLSEA 179
QY 179 DVRALGLACDLPGRFVAESAEEVLLPRLVSCPGPLDODQOEAARALOGGPPYPSTW 238
Db 181 DVRALGLACDLPGRFVAESAEEVLLPRLVSCPGPLDODQOEAARALOGGPPYPSTW 240

```

```

Db 180 DVALGGLACLDLPGRFVAESAIEVLLPRLVSCPGPLDQDOEAAARALOGGGPPGPSTW 239
Qy 239 SVSTMDALRGILLPVLGOPITISIPQGIYAAMROSSRPSMRPOPERTLLRRFRREVEXT 298
Db 240 SVSTMDALRGILLPVLGOPITISIPQGIYAAMROSSRPSMRPOPERTLLRRFRREVEXT 299
Qy 299 ACPGSKKAREIDESLIFPKKWELEACVDAALLATQMDVNAIPFEYEDLDVKKHLDLY 358
Db 300 ACPGSKKAREIDESLIFPKKWELEACVDAALLATQMDVNAIPFEYEDLDVKKHLDLY 359
Qy 359 PGQPESVYIOMLGYFLFKMSPEDIRKMNVTSLKALLEVDKGHEMSPOAPRRPLPOVA 418
Db 360 PGQPESVYIOMLGYFLFKMSPEDIRKMNVTSLKALLEVDKGHEMSPOAPRRPLPOVA 411
Qy 419 TLIDRFVKGROLDKDTLDTLATFYPGYLCSLSPBELSSVPPSSITWAVRPDOLDTCDRQ 478
Db 412 TLIDRFVKGROLDKDTLDTLATFYPGYLCSLSPBELSSVPPSSITWAVRPDOLDTCDRQ 471
Qy 479 LDVLYPKARLAFQNMNGSEYFVKIOSFLGAPTEBDLKAISOONVSMDLATFMKLTDAVL 538
Db 472 LDVLYPKARLAFQNMNGSEYFVKIOSFLGAPTEBDLKAISOONVSMDLATFMKLTDAVL 531
Qy 539 PLTVAEVOKLLGPHVEGLKAEBERHPRVDWILRQRODDLTGLGLOGGIPNGYLVLDLS 598
Db 532 PLTVAEVOKLLGPHVEGLKAEBERHPRVDWILRQRODDLTGLGLOGGIPNGYLVLDLS 591
Qy 599 VOETLSGTPCLLPGPVLTVALLASTLA 628
Db 592 VOETLSGTPCLLPGPVLTVALLASTLA 621

```

```

RESULT 4
Q9ERA7 PRELIMINARY: PRT: 625 AA.
ID Q9ERA7
AC Q9ERA7
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE MESOTHELIN.
GN ERG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EKER RAT.
RX MEDLINE=20403886; PubMed=10944454;
RA Yamashita Y., Yokoyama M., Kobayashi E., Takai S., Hino O.;
RT "Mapping and determination of the cDNA sequence of the Erc gene
RT preferentially expressed in renal cell carcinoma in the Tsc2 gene
RT mutant (Eker) rat model."
RL Biochem. Biophys. Res. Commun. 275:134-140(2000).
DR EMBL; D87351; BAB13512.1;
SQ SEQUENCE 625 AA; 68852 MW; 343EF77ADF390CD2 CRC64;

```

Query Match 52.4%; Score 1710; DB 11; Length 625;  
 Best Local Similarity 55.9%; Pred. No. 8.2e-123;  
 Matches 352; Conservative 80; Mismatches 184; Indels 14; Gaps 5;

```

Qy 1 MALQRLDP-CWSCGDR--PGSLFLFLSLGWNVPARTLAGGTESAPLGVLTTPHNIS 57
Db 1 MALPAPRLDGLSCGSPICSRSLILLSLGWLPRLOQTOTRTTSQDAALLHNAVTV-DFA 59
Qy 58 SLSPROLLGFCPEAVSGLSTERVRELAVAAOKNVKLTSEDLRCLAHRISE--PPEDIDA 115
Db 60 SLPTGLTGLTCDCEVSGLSMGHAKELAAVAVROKNIVLQVHRCRLARLPKLTNEEDA 119
Qy 116 LPDLTLFLNPAESGPOACTFEFSRITKANVDLPRGAREQRLLPALACWYRGSL 175
Db 120 LPDLTLFLNPAEPGQACAHFSLISKANVNVLPRLSLEROLLTGALCKGQVYGFQV 179

```

```

Qy 176 SEADVRLGGLACLDLPGRFVAESAIEVLLPRLVSCPGPLDQDOEAAARALOGGGPPGP 235
Db 180 SETDARALGGLACLDLPGRFVAKSSSEVLLPMLARCCGPLDQDOAKAVREVLSGARPIGP 239
Qy 236 STWSVSTMDALRGILLPVLGOPITISIPQGIYAAMROSSRPSMRPOPERTLLRRFRREV 295
Db 240 STWSVSTMDALRGILLPVLDSEIVSHIPKDVITEMLOGISRPSRLGSKWTVTHPRFRDT 299
Qy 296 EKTACPSKAREIDESLIFPKKWELEACVDAALLATQMDVNAIPFEYEDLDVKKHLD 355
Db 300 EOKACPPEKEPNVDNLSLIFQNMWELEACVDTLACGMDLVNELIPFEYEDLSIFKHL 359
Qy 356 ELYPGPESVYIOMLGYFLFKMSPEDIRKMNVTSLKALLEVDKGHEMSPOAPRRPLP 415
Db 360 KTYPGYVESLITKOGHFRFVYSPEDIRQMNVTSPDVTNTLTKVSKGKMD-----A 411
Qy 416 QVATLIDRFVKGROLDKDTLDTLATFYPGYLCSLSPBELSSVPPSSITWAVRPDOLDTC 475
Db 412 QVATLIDRFVKGROLDKDTLDTLATFYPGYLCSLSPBELSSVPPSSITWAVRPDOLDTC 471
Qy 476 PROLDVLYPKARLAFQNMNGSEYFVKIOSFLGAPTEBDLKAISOONVSMDLATFMKLT 535
Db 472 ORHGLITLQKACSAFQNVSGLEYFEKIRTFVGASREDLRALSOONVSMDLATFMKLT 531
Qy 536 ALPLTVAEVOKLLGPHVEGLKAEBERHPRVDWILRQRODDLTGLGLOGGIPNGYLV 595
Db 532 ALVGLSVAEVOKLLGPHIGDKLTEDKSPVDRWILFRQOKNDLSGLGLOGGIPNGYLV 591
Qy 596 DLSTOETLSGTPCLLPGPVLTVALLASTLA 625
Db 592 DFNVEAFSSGAPLIGPGVFAMIPALLSA 621

```

```

RESULT 5
Q61468 PRELIMINARY: PRT: 625 AA.
ID Q61468
AC Q61468
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE MEGAKARYOCYTE POTENTIATING FACTOR PRECURSOR.
GN MSLN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kojima T., Taniguchi Y., Hattori K., Oh-eda M.;
RT "mouse Megakaryocyte Potentiating Factor cDNA";
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95394969; PubMed=7665620;
RA Kojima T., Oh-eda M., Hattori K., Taniguchi Y., Tamura M., Ochi N.,
RA Yamaguchi N.;
RT "Molecular cloning and expression of megakaryocyte potentiating factor
RT cDNA."
RL J. Biol. Chem. 270:21984-21990(1995).
DR EMBL; D86370; BAA13077.1;
DR MCD; MGI:1888992; Msln.
FT SIGNAL.
KW SIGNAL.
SQ SEQUENCE 625 AA; 69423 MW; 9A5E9C3222C6983F CRC64;

```

Query Match 52.3%; Score 1707; DB 11; Length 625;  
 Best Local Similarity 56.1%; Pred. No. 1.4e-122;  
 Matches 352; Conservative 80; Mismatches 182; Indels 14; Gaps 5;

```

Qy 1 MALQRLDP-CWSCGDR--PGSLFLFLSLGWNVPARTLAGGTESAPLGVLTTPHNIS 57
Db 1 MALPAPRLDGLSCGSPICSRSLILLSLGWLPRLOQTOTRTTSQDAALLHNAV-NGAADFA 59

```

```

QY 58 SLSPROLLGFPCAESVSGSTEREVELAVALAOKNVLSTEQRLCLAHRLSE--PEDDLA 115
DB 60 SLPTGLFGLTCEEVSDLSMEQAKGLAMAVRQKNITLNGHQRLCLARRLPHRLDDEELNA 119
QY 116 LPDLLELFLNPDASEGPOACTRFESRITKANVDLPGCAPERQRLPALACWYRGSLT 175
DB 120 LPDLLELFLNPMFPGQOACAFHFFSLISKANDVLPRLSRLEORLLMEALKOCGYGGEV 179
QY 176 SEADRALGGLACDLPGRFVAESAELLPLVSCBPLDDOQEAARALOGGCGPPYGP 235
DB 180 SEADRALGGLACDLPGRFVARSSVLLPWLACGCPPLDOSEKAVREYLRSGRTQYGP 239
QY 236 STWVSYTDALRGLLPVLGPIIRSIPOGIYAAMRORSSRDPSPWROPERTILRPRFRREV 295
DB 240 SKMSYSTDALQSLVAVDESIVQSIIPKOVAKEMLOHISRDSRSGSKTIVHPFRRA 299
QY 266 EKTAPSGKKAREIDESLIFTKKMLEACVDALLATQMDRVNAIPFTYEQLDVLKHKID 355
DB 300 BOKACPPGKEPKYDEDELIIFYONMELACVCGTMLAROMDLVNEIPFTYEQLSIFKHID 359
QY 356 ELYPOGYPESVIOHLYFLKMSPEIDIRKMNVTSLLETALKLEVDKGHEMSPOARRPLP 415
DB 360 KTYPOGTPESLIOQLGHEFFRIVSPEDIHOMNVTSPDTVKTLLKVKSGQKMAQA----- 413
QY 416 QVATLIDRFVKGROLDKDTLDTLTAFFPGYLCSPBELSSVPPSSIMAVRPODLDTCD 475
DB 414 --IALVACYLRRGGQLDEDMVKALGDIPLSYLCDFSPDLHSPSSVMVLVGPDLKDKS 471
QY 476 PROLDVLYPKARLAFQNNNGSEYFVKIOSFLGAPTEDLKALSOONVSMDLATPMKLRID 535
DB 472 ORHGLGLYOKACSAFQNNNGSEYFVKIOSFLGASVCKDLRALSOHNVSMDLITFKRLQVD 531
QY 536 AVLPFLTVAEVOKLLGPHVEGLKAEERHPRVDMILRORODDLDTLGLGLOGGIPNGYLV 595
DB 532 SLVGSVAEVOKLLGPNVVDLKTEDKSPVRDMLFRQHKDLRLGLGLOGGIPNGYLV 591
QY 596 DLVSQETLSGTPCLLPGPVLTVLALL 623
DB 592 DENVREAFSSRASLLGPGFVLWIPALL 619

RESULT 6
QY 09UK57 PRELIMINARY; PRT; 398 AA.
AC Q9UK57;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MESOTHELIN/MEGAKARYOCYTE POTENTIATING FACTOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Scholler N., Fu N., Yang Y., Ye Z., Goodman G.E., Hellstroem K.E.,
RA Hellstroem I.;
RT "Soluble member(s) of the mesothelin/megakaryocyte potentiating factor
RT family are detectable in sera from patients with ovarian carcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 19:11531-11536(1999).
DR EMBL; AF180951; AAF01409.1; -
DR InterPro; IPR000585; Hemoexin.
DR PROSITE; PS00024; HEMOEXIN; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 398
FT SEQUENCE 398 AA; 43811 MW; 2D542C407807C498 CRC64;

```

Query Match 46.8%; Score 1526; DB 4; Length 398;  
 Best Local Similarity 97.1%; Pred. No. 5.6e-109;  
 Matches 297; Conservative 1; Mismatches 0; Indels 8; Gaps 1;

```

QY 295 VEKTACPSGKKAREIDESLIFYKKMLEACVDALLATQMDRVNAIPFTYEQLDVLKHKL 354
DB 1 VEKTACPSGKKAREIDESLIFYKKMLEACVDALLATQMDRVNAIPFTYEQLDVLKHKL 60
QY 355 DELYPOGYPESVIOHLYFLKMSPEIDIRKMNVTSLLETALKLEVDKGHEMSPOARRPL 414
DB 61 DELYPOGYPESVIOHLYFLKMSPEIDIRKMNVTSLLETALKLEVDKGHEMS----- 112
QY 415 POVATLIDRFVKGROLDKDTLDTLTAFFPGYLCSPBELSSVPPSSIMAVRPODLDT 474
DB 113 POVATLIDRFVKGROLDKDTLDTLTAFFPGYLCSPBELSSVPPSSIMAVRPODLDT 172
QY 475 DPROLDVLYPKARLAFQNNNGSEYFVKIOSFLGAPTEDLKALSOONVSMDLATPMKLR 534
DB 173 DPROLDVLYPKARLAFQNNNGSEYFVKIOSFLGAPTEDLKALSOONVSMDLATPMKLR 232
QY 536 DAVLPFLTVAEVOKLLGPHVEGLKAEERHPRVDMILRORODDLDTLGLGLOGGIPNGYLV 594
DB 233 DAVLPFLTVAEVOKLLGPHVEGLKAEERHPRVDMILRORODDLDTLGLGLOGGIPNGYLV 292
QY 595 LDLSVQ 600
DB 293 LDLSVQ 298

RESULT 7
QY 09FLJ5 PRELIMINARY; PRT; 698 AA.
AC 09FLJ5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PUTATIVE HRPO.
GN PUTATIVE HRPO.
OS Pseudomonas glumae.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=337;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-GM4;
RA Takikawa Y., Eo S., Adachi S., Kojima M.;
RT "Deduced hrp genes from Burkholderia glumae and Acidovorax avenae
RT subsp. avenae.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB053455; BAB20913.1; -
DR InterPro; IPR001712; Bact_exp_FHPRP.
DR Pfam; PF00771; FHPRP; 1.
DR PRINTS; PR00949; TYPE3IMAPROT.
FT SEQUENCE 698 AA; 75549 MW; 4B8BE407CE559A3 CRC64;

```

Query Match 4.1%; Score 133; DB 2; Length 698;  
 Best Local Similarity 23.5%; Pred. No. 0.11; Mismatches 209; Indels 160; Gaps 31;  
 Matches 135; Conservative 70; Mismatches 209; Indels 160; Gaps 31;

```

QY 37 GETGTESAPLGGVLTTPHNISLSPROLLGPCAESVSGSTEREVELAVALAOKNVLST 96
DB 218 GMSADDAANRPSVLSVGDAWSQIRSL-----SVAAGVMTIRVADENQA---KRSGLD 270
QY 97 EQLRLCLARHLSPPEDLDALPLDLLF-LNPD-----AFSG-----POAC 135
DB 271 E-----IGHQLGSSRALFFFAAVILLAFALVGFPSLLFVLLAAALSFAGYRLSVKPPSR 326
QY 136 TRFESRITKANVDLPGAPERQRLPAA--LAC-WGYRGSLSTADYRALGGLACDLP 192
DB 327 DRAMRRETLASQRYGR-KTEVPPLIPRAPDFACAVGR-----ISPDL-- 369
QY 193 RFVAESAELLPLVSCBPLDDOQEAARALOGG-GPPYGPSTW-----SVSTMDL 246
DB 370 -----AANVAAPTLDA-----FEAKRALDQALALPPPGITMWNAHPVATTEIL 416
QY 247 RGLLPVL-----GQ--PIIRSIPOGIYAAMRORSSRDPSPWROPERTILRPRFRREV 297

```

```

Db 417 IHDPRLHMTWPPCKAMLPDLRLVPAEVA--SARSCWMPRR-----RRRPVDS 466
QY 298 TACS-GKKAREIDESLIFYKKELE-----ACVDALLATQMDRVNAIPFTTEOL 347
Db 467 AGFRLRLRRAGAGADAV---NRTEOIIAHLVAPGATATFLGIGIOEYOMI-----L 516
QY 348 DVLHKHKLDELPGYRPSVIOHLYGLFKMSPEDIRKMNWTS-LETKALLEVKGHEMS 406
Db 517 DQSHDAPGLVAEYQKVLPPORADVLRRLLEBOVSIRNWSFESLWACAKEDMLML 576
QY 407 POAERRLPOVATLIDREVKRGQDLKDTLDTLTAFFPGYLSLSPPE-----LSVPPS 461
Db 577 TEYVRGDLSPF--LAHNAKGERQLS-----AVLFQMAQGHIRAKIKQPTG 622
QY 462 SIMAVRQDLDTCPPROLDVLYPKARLAFONMNGSEYFKIQSELGAPTEDKALSQON 521
Db 623 NFLALPPND-----NAF-----LIDRIQSEVGAAPREGVVLVT--- 655
QY 522 VSMDLATEFMKLRDVLPL-LTFAEYOKLSPHE 554
Db 656 -SMDIRRYVRMIESRLGMLSVYSIOE-LGEHVE 687

```

```

RESULT 8
ID 09UGS3 PRELIMINARY: PRT: 797 AA.
AC 09UGS3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE D1756623.1 (NOVEL LEUCINE RICH PROTEIN) (FRAGMENT).
GN D1756623.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eulheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RA Corby N.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035681; CAB63072.1;
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR000483; LRR_Cterm.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003591; LRR_Lyp.
DR InterPro: IPR003592; LRR_Cut.
DR Pfam; PF00560; LRR; 18.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00370; LRR; 8.
DR SMART; SM00082; LRRCT; 2.
DR SMART; SM00013; LRRNT; 2.
DR SMART; SM00369; LRR_Typ; 4.
FT NON_TER 797
SQ SEQUENCE 797 AA; 85614 MW; 8C3247883EAE59AD CRC64;

```

```

Query Match 4.0%; Score 131.5; DB 4; Length 797;
Best Local Similarity 24.4%; Pred. No. 0.17;
Matches 113; Conservative 45; Mismatches 181; Indels 125; Gaps 23;
QY 15 RPSGL--LFLFSGWHPARTLAGETTESAPLGGLVITTHN-ISSLSPPQLIGFPCAE 71
Db 191 RRGFGALGALATLNLAINALVYLPAMAFQDLRLVRLSLSHMALSVLAPALGLPALR 250
QY 72 VSGSTERVRELAAVALAKNVKLTSTEOURLAH-RLSEPP-----EDLDLPL--LDLL 122
Db 251 RLSLHNLQALPGPV-----LSQARGLARLELGNPLITYAGEEGCLALPGIRELL 302
QY 123 -----FLNPDAFSGPOACTRFFSRITKAN--VDLLP---RGAPERQRLPALAAC--- 167

```

```

Db 303 DGALQALGPRAFA---HCPRLHTLDLGNQDLDTPLPGGPGQLRLRLQGNPLMCCQA 359
QY 168 -----WVGRLSLSEADYRALGGLADLPGRFVAESAENVLLPRLVSCPGPLDQOQEAR 222
Db 360 RPLLEW-----LARARSDG--ACQPRRLRGEALDALRPMQLRCPGDAAQEEELIE 411
QY 223 AALQG-----GPPYGPSTWSVSTMDLRGLPVLGPI-----ISIPQIVAAWROR 272
Db 412 RAVAGPRAPRPPRGGEERAVALPCPRACVCYVESHSSEGGGLQAVPRGF----- 464
QY 273 SSRDPSWRQPERTILRPFRREVEKTACPSGKAREIDESLIFYKKELEACVDALLAT 332
Db 465 -----PS--DTQLDLRLRNHFPSPVRAFP-----GLGHLVSLHLCGIAELEAG 508
QY 333 QMDRVNAIPFTY---EQLDVLRHKLDELXPGYRPSVIOHLYGLFKMSPEDIRKMNWTS 389
Db 509 ALAGLRLITYLSDNQLAGLSAALLEGAPF-----LGYLYLEKN----- 548
QY 390 LETKALLEVDKGHEMSPOAERRLPOVATL-----IDREVKG 427
Db 549 -----RFLQV-----PGALRALPLSLHLQDNNAVDRILAPG 580

```

```

RESULT 9
ID 09WXJ0 PRELIMINARY: PRT: 939 AA.
AC 09WXJ0;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ACONITASE (FRAGMENT).
GN ACON.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=1718;
RN 11
RP SEQUENCE FROM N.A.
RA Nakamura J., Kimura E., Oosumi T., Nakatsu T.;
RT "Brevibacterium lactofermentum ATCC 13639 acn gene for Aconitase.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB025424; BAA76717.1;
DR HSSP; P16276; 180J.
DR InterPro: IPR001030; Aconitase.
DR InterPro: IPR000573; Aconitase_C.
DR Pfam; PF00330; aconitase; 2.
DR Pfam; PF00694; Aconitase_C; 1.
DR PRINTS; PR00415; ACONITASE.
DR PRODOM; PD000511; Aconitase; 1.
DR PROSITE; PS00450; ACONITASE_1; UNKNOWN_1.
DR PROSITE; PS01244; ACONITASE_2; 1.
FT NON_TER 1
SQ SEQUENCE 939 AA; 101771 MW; E2DE6472D7F4A096 CRC64;

```

```

Query Match 3.9%; Score 126; DB 2; Length 939;
Best Local Similarity 20.3%; Pred. No. 0.56;
Matches 137; Conservative 76; Mismatches 205; Indels 258; Gaps 35;
QY 60 SPROLL-----GPCAESGSLSTERVRELAAVALAKNVKLTSTEOURLCLAHRLSEPEDLD 114
Db 79 TPARYLMDFTGVC--VVDLAT--MRAVAA-----LGQDPNDVN 115
QY 115 AL-PLDLL--FLNPDAFSGPOACTRFFSRITKANVDLLPAGAPERQRLPALAACGVR 171
Db 116 PLNPDAEVYIDHSYIVEAFGRPDALAK-----NVEIEYRNEERYQFLR-----WCSE 162
QY 172 GSLSSEADYRALGGLADLPGRFVAESAENVLLPRLVSCPGPLDQOQEAALAQGGPP 231
Db 163 -----SFSNFRVVP-----PQTGIVHOVN-----ETLAVVDENSLA 196
QY 232 YGPSTW-----SVSTMDLRGLL-----PVLGQPIIRSIPOGIIVAAWRQSSRD 276

```







RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.,  
 RT "Sequence and analysis of chromosome II of *Arabidopsis thaliana*."  
 RL Nature 402:761-768(1999).  
 DR EMBL: AC006202; AAD29825.1;  
 SO SEQUENCE 2218 AA; 235758 MW; 525B6B1937A62998 CRC64;

Query Match 3.7%; Score 120; DB 10; Length 2218;  
 Best Local Similarity 19.1%; Pred. No. 5.7;  
 Matches 135; Conservative 81; Mismatches 222; Indels 270; Gaps 32;

QY 30 HPARTLA-----GETGESAPLGGVLTTPHNISLSPPROLLGFPACAEVSGLSTERE 82  
 Db 169 HPTSSLATSPDLGPPGFQSLP-----ASPAPTPIRGGRGRSGRGAAGRR 217  
 QY 83 LAVALAQNKLVS--TEQRLCLAHRLSEPPEDLDLPLDLLFLNPDASFSGQACTRFFS 140  
 Db 218 VEGVLHGSNNSITORTETATSLA-----SDAEA-----TFALPRASSELVS 259  
 QY 141 RITKAMVDLPRGAPER-ORLLPALACGWYRGSILSEADYRVALGGLCDLPGFVAESA 199  
 Db 260 RVPKAN-----EGSTSNPDQVSPVHSATTALRSOKAKDKL-----DAPPGDSGSH 306  
 QY 200 EYLLPRLYSCGPIDDQOQEA-----RAALQGGGPPYGPSTMSV-----STMD 244  
 Db 307 -----VQTLNVLNLSSEKRAFAVKRRPLIOGGGPVONNAVSSVCGSKSPSEGRY 359  
 QY 245 ALRGL-----LPVLGQPI-----IRSIPOGIYAAMQR--- 272  
 Db 360 ALQGVTTAPSAITLPMSSQPSDAILPMSSQVGTVEAOEANVPSLPALPAKRVRNLP 419  
 QY 273 SSRPSWRQPER-----TIL-----RPRFR 293  
 Db 420 SRGETPKRQGRGQPLPATDASSARSTGLTPQLEVKVGNLSGTAKPKDAYAKQPHFSQ 479  
 QY 294 EYERTACSGSKKAREIDSLIFKKWELEACVDALLATQMDRNAIFPTTEQLDVLYKH 353  
 Db 480 SVAPDIHSSGSLSOEIRDTSGTGSARKQADYTDVARVKET-----FSESLKHK 533  
 QY 354 LDELYPQGPESVIOHGLYFLKMSPEDIRKMNVTSLTLK-----ALLEVKG 402  
 Db 534 VGE--PSATTTTNVPD-----AOSPGEM--NLHTVETHKAEDESSGLKNOBALYNLSKA 582  
 QY 403 HEMSPQARRPLPOVATL-----TAIFYGYLCSLSPELSSV-----PPSITMAVR 467  
 Db 583 DKIVSDIP-HVPDGLTTSVGVANKVDIGSSKVAENELVKIPGVDVSSVIOISLQNT 641  
 QY 425 VKRGQLDKDTLDL-----TAIFYGYLCSLSPELSSV-----PPSITMAVR 467  
 Db 642 LTAKSSLEKCTADOLGKLSQEGETTPASDGETCHLAETFASSLYVRSSEPTASA--- 697  
 QY 468 PQDDTQOPRQDLVLYPKARLAFON---MNGSEFYVKIQSEFGAPREDLKALSOQNV 523  
 Db 698 ---STTAEPDLPTDKL--EKNISFODEVKTLLGD---KREAIL--LSSEQITNVN 741  
 QY 524 MDLAT---FMKLRDVAVLPTVAEVQKLGPHEGLKAERHRPVRD 567  
 Db 742 SKIETNSELOASRTDEV-----PHYDGKSVDVANQTVKE 776

RESULT 14  
 ID 09AUB4 PRELIMINARY; PRT; 3574 AA.  
 AC 09AUB4;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DE PUTATIVE CHROMATIN REMODELING PROTEIN SYD.  
 GN SPLAYED.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA Wagner D., Meyerowitz E.M.;  
 RT "SPLAYED, a Putative Chromatin Remodeling Factor, Controls  
 RT Reproductive Development in Arabidopsis."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF247809; AAK31908.1;  
 SO SEQUENCE 3574 AA; 389826 MW; 0CA25C1FLIAB6A0 CRC64;

Query Match 3.6%; Score 117.5; DB 10; Length 3574;  
 Best Local Similarity 19.2%; Pred. No. 18;  
 Matches 138; Conservative 82; Mismatches 233; Indels 265; Gaps 32;

QY 30 HPARTLA-----GETGESAPLGGVLTTPHNISLSPPROLLGFPACAEVSGLSTERE 82  
 Db 1500 HPTSSLATSPDLGPPGFQSLP-----ASPAPTPIRGGRGRSGRGAAGRR 1548  
 QY 83 LAVALAQNKLVS--TEQRLCLAHRLSEPPEDLDLPLDLLFLNPDASFSGQACTRFFS 140  
 Db 1549 VEGVLHGSNNSITORTETATSLA-----SDAEA-----TFALPRASSELVS 1590  
 QY 141 RITKAMVDLPRGAPER-ORLLPALACGWYRGSILSEAD-----VRALGGLA 187  
 Db 1591 RVPKAN-----EGSTSNPDQVSPVHSATTALRSOKAKDKLDAPPGDSGHVOTLNVLE 1645  
 QY 188 CDLPGRFAESAENVLLPRLVS--CGEPILDO--DQCEAANAALQGGGPPYGPSTMSV--- 240  
 Db 1646 NSSERKARFAVKRRPLIOGVSSOHGPNKOPDLPLVSTSTLLGGGPVONNAVSSVCGS 1705  
 QY 241 ---STMDALRGL-----LPVLGQPI-----IRSIPOGI 265  
 Db 1706 KSPSGRITYTLQGVTTAPSDAILPMSSQVGTVEAOEANVPSLPAL 1765  
 QY 266 VAAWRQ---SSRPSWRQPER-----TIL-----RPRFR 287  
 Db 1766 PAKRRVRNLP-SRGETPKRQGRGQPLPATDASSARSTGLTPQLEVKVGNLSGTAKPKDA 1825  
 QY 288 ---RPRFRREYERKACSGSKKAREIDSLIFKKWELEACVDALLATQMDRNAIFPT 343  
 Db 1826 VAKQEPHFSSQVAPDIHSSGSLSOEIRDTSGTGSARKQADYTDVARVKET----- 1879  
 QY 344 YEQDLVLYKHKDELYPQGPESVIOHGLYFLKMSPEDIRKMNVTSLTLK----- 394  
 Db 1880 FSEISLKHKGE--PSATTTTNVPD-----AOSPGEM--NLHTVETHKAEDESSGLKN 1928  
 QY 395 --ALLEVDKHEMSQARRPLPOVATL-----TAIFYGYLCSLSPELSSV----- 420  
 Db 1929 QEALYNLKSKDKIVSDIP-HVPDGLTTSVGVANKVDIGSSKVAENELVKIPGVDVS 1987  
 QY 421 ---IDRFYKRGQLDKDTLDL-----TAIFYGYLCSLSPELSSV----- 458  
 Db 1988 SVIOLSLQNTLTAKSSLEKCTADOLGKLSQEGETTPASDGETCHLAETFASSLSYVS 2047  
 QY 459 -PPSITMAVRPQDDTQOPRQDLVLYPKARLAFON---MNGSEFYVKIQSEFGAPTED 513  
 Db 2048 EPTASA-----STTAEPDLPTDKL--EKNISFODEVKTLLGD---KREAIL----- 2087  
 QY 514 LKALSOQNVSMDLAT---FMKLRDVAVLPTVAEVQKLGPHEGLKAERHRPVRD 567  
 Db 2088 LSSEQITNVNKSIEINSELOASRTDEV-----PHYDGKSVDVANQTVKE 2132

RESULT 15  
 ID 09U5Y1 PRELIMINARY; PRT; 2015 AA.  
 AC 09U5Y1;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-JUN-2001 (TREMblrel. 13, Last sequence update)  
 DE MICROTUBULE-ASSOCIATED PROTEIN CP224.

OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX2;  
 RA Graef R., Daudeker C., Schliwa M.;  
 RT "Dictyostelium Ddcp224 is a microtubule-associated protein and a  
 permanent centrosomal resident involved in centrosome duplication.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ012088; CAB56504.1;  
 DR InterPro: IPR000357; HEAT\_repeat.  
 DR PROSITE: PS50077; HEAT\_REPEAT; 1.  
 SO SEQUENCE 2015 AA; 224038 MW; BA64E982ADDC92EE CRC64;

Query Match 3.68; Score 116.5; DB 5; Length 2015;  
 Best Local Similarity 23.3%; Pred. No. 9.2;  
 Matches 66; Conservative 30; Mismatches 74; Indels 113; Gaps 13;

```

QY 355 DELYPGYPESVIGHLGYLFILKMSPEDIRKKNVTSLET----- 393
Db 3 DEAPSGSIEDRIHKKM-----KRVSGLEELTKFRNSIEGSGPLFNEWGPQ 51
QY 394 -KALLEVDKGHEMSPOAPRRPLPOVATLIDRFVKGRLDKDLDLTAFYPGYL----- 447
Db 52 FKKILA-----DINPMGOERALEPLSAFIDRC-----DCVNKFAASYVGVLYEKL 97
QY 448 -----CSLSPPELSSVPP-----SSIWAVRPDLDTCDP 476
Db 98 ASTRPRAKEKTIIECLLTLEADSAEPVVEALIKGTSTSPKILLASL-AALTOALKTFGP 156
QY 477 ROLPVLPRKARLA--FONMN-----GSEYFVKIOSFLGA-----PTEDLKALS 518
Db 157 KQIVYKLLIKQFSPWFENRDKGIRDQASSELFETIYRWIGKALILPILISEALTPIQ-LKALQ 215
QY 519 QQNVSMDLATFMKLRTPDAVLPL--TVAEVOKLIGPHVEGLKAE 559
Db 216 DQ-----FEKLPPTPAVPLKTYRSEAAKALANASKGIOAK 250

```

Search completed: December 7, 2001, 09:53:43  
 Job time: 326 sec

**THIS PAGE BLANK (USPTO)**